

THE INVERSE CONJECTURE FOR THE GOWERS NORM OVER FINITE FIELDS IN LOW CHARACTERISTIC

TERENCE TAO AND TAMAR ZIEGLER

ABSTRACT. We establish the *inverse conjecture for the Gowers norm over finite fields*, which asserts (roughly speaking) that if a bounded function $f : V \rightarrow \mathbb{C}$ on a finite-dimensional vector space V over a finite field \mathbb{F} has large Gowers uniformity norm $\|f\|_{U^{s+1}(V)}$, then there exists a (non-classical) polynomial $P : V \rightarrow \mathbb{T}$ of degree at most s such that f correlates with the phase $e(P) = e^{2\pi i P}$. This conjecture had already been established in the “high characteristic case”, when the characteristic of \mathbb{F} is at least as large as s . Our proof relies on the weak form of the inverse conjecture established earlier by the authors and Bergelson [3], together with new results on the structure and equidistribution of non-classical polynomials, in the spirit of the work of Green and the first author [22] and of Kaufman and Lovett [28].

1. INTRODUCTION

1.1. The inverse conjecture. Let $\mathbb{F} = \mathbb{F}_p$ be a finite field of prime order $\text{char}(\mathbb{F}) = p$. Throughout this paper, \mathbb{F} will be considered fixed (e.g. $\mathbb{F} = \mathbb{F}_2$ or $\mathbb{F} = \mathbb{F}_3$), and the term “vector space” will be shorthand for “vector space over \mathbb{F} ”, and more generally any linear algebra term (e.g. span, independence, basis, subspace, linear transformation, etc.) will be understood to be over the field \mathbb{F} unless otherwise stated.

If V is a vector space, $f : V \rightarrow \mathbb{C}$ is a function, and $h \in V$ is a shift, we define the *multiplicative derivative* $\Delta_h f : V \rightarrow \mathbb{C}$ of f by the formula

$$\Delta_h f := (T_h f) \overline{f}$$

where the shift operator T_h with shift h is defined by $T_h f(x) := f(x + h)$. If V is finite, and $d \geq 1$ is an integer, we define the *Gowers uniformity norm* $\|f\|_{U^d(V)}$ by the formula

$$\|f\|_{U^d(V)} := |\mathbb{E}_{h_1, \dots, h_d, n \in V} \Delta_{h_1} \dots \Delta_{h_d} f(n)|^{1/2^d},$$

where we use the expectation notation $\mathbb{E}_{a \in A} f(a) := \frac{1}{|A|} \sum_{a \in A} f(a)$ for any finite non-empty set A , with $|A|$ denoting the cardinality of A . We review some basic properties of the Gowers uniformity norms in Appendix B.

The *inverse conjecture* for the Gowers norm in finite characteristic addresses the question of determining those bounded functions $f : V \rightarrow \mathbb{C}$ with large Gowers norm. To

1991 *Mathematics Subject Classification.* 11B30, 11T06.

The first author is supported by a grant from the MacArthur Foundation, and by NSF grant CCF-0649473. The second author is supported by ISF grant 557/08, and by an Alon fellowship.

phrase this conjecture correctly in the low characteristic case, we need the notion of a *non-classical polynomial*¹:

Definition 1.2 (Polynomials). Let V be a finite-dimensional vector space, let $d \geq 0$ be an integer, and let G be an additive group. A function $P : V \rightarrow G$ from V to G is said to be a (*non-classical*) *polynomial* of degree $\leq d$ if one has

$$\Delta_{h_1} \dots \Delta_{h_{d+1}} f(n) = 0$$

for all $h_1, \dots, h_{d+1}, n \in V$, where $\Delta_h f(n) := (T_h - 1)f(n) = f(n+h) - f(n)$ is the additive derivative of f in the direction h . We adopt the convention that the zero polynomial 0 has degree $-\infty$. We denote the space of all polynomials of degree $\leq d$ as $\text{Poly}_{\leq d}(V \rightarrow G)$; this is clearly an additive group, with $\text{Poly}_{\leq d}(V \rightarrow G) = \{0\}$ for $d < 0$.

Remark 1.3. In practice the group G will usually be the unit circle $\mathbb{T} := \mathbb{R}/\mathbb{Z}$, the finite field \mathbb{F} , or the embedded copy $\iota(\mathbb{F})$ of \mathbb{F} in \mathbb{T} , where $\iota : \mathbb{F} \rightarrow \mathbb{T}$ is the additive homomorphism

$$\iota(j) := \frac{j}{p} \bmod 1.$$

In particular, $\iota(\mathbb{F}) = \frac{1}{p}\mathbb{Z}/\mathbb{Z}$ is the group of p^{th} roots of unity in \mathbb{R}/\mathbb{Z} .

Remark 1.4. Polynomials that take values in \mathbb{F} (and by abuse of notation, $\iota(\mathbb{F})$) will be referred to as *classical polynomials*; but in this paper, the term “polynomial” will be understood to encompass the non-classical case unless otherwise stated. Clearly ι induces an isomorphism $\iota_* : \text{Poly}_{\leq d}(V \rightarrow \mathbb{F}) \rightarrow \text{Poly}_{\leq d}(V \rightarrow \iota(\mathbb{F}))$. We will take advantage of this isomorphism whenever we need to use the multiplicative structure on \mathbb{F} , since \mathbb{T} has no multiplicative structure, save for the fact that it is a \mathbb{Z} -module (i.e. one can define $n\alpha$ when $\alpha \in \mathbb{T}$ and $n \in \mathbb{Z}$).

Remark 1.5. These notions of polynomials are part of a larger theory of *polynomial algebra* between (filtered) groups that are not necessarily abelian; see Appendix C.

Example 1.6. The map $P : \mathbb{F}_2 \rightarrow \mathbb{T}$ with $P(0) := 0$ and $P(1) := 1/2$ is a classical polynomial of degree ≤ 1 . The map $Q : \mathbb{F}_2 \rightarrow \mathbb{T}$ with $Q(0) = 0$ and $Q(1) = 1/4$ is a (non-classical) polynomial of degree ≤ 2 ; note that $\Delta_1 Q = \frac{1}{4} - P$.

We can generalise these examples to higher dimensions. If $n \geq 1$ is an integer, we let $L : \mathbb{F}_2^n \rightarrow \mathbb{Z}$ be the function $L(x_1, \dots, x_n) := |x_1| + \dots + |x_n|$, where $x \mapsto |x|$ is the obvious map from \mathbb{F}_2 to the fundamental domain $\{0, 1\}$. This map is not a polynomial (either in the classical or non-classical sense); however, the function $\frac{L}{2} \bmod 1$ is a classical polynomial of degree ≤ 1 from \mathbb{F}_2^n to \mathbb{T} , $\frac{L}{4} \bmod 1$ is a (non-classical) polynomial of degree ≤ 2 from \mathbb{F}_2^n to \mathbb{T} , and more generally, for any $k \geq 0$, $\frac{L}{2^{k+1}} \bmod 1$ is a (non-classical) polynomial of degree $\leq k$ from \mathbb{F}_2^n to \mathbb{T} . For further generalisation of these examples, see Lemma 1.7(ii), (iii) below.

The relevance of (non-classical) polynomials to the Gowers norms can be seen from the easily verified fact that if $f : V \rightarrow \mathbb{C}$ is a function on a finite-dimensional vector space

¹Strictly speaking, “not necessarily classical polynomial” would be a more accurate terminology than “non-classical polynomial”.

V that is bounded in magnitude by 1 (thus $|f(x)| \leq 1$ for all $x \in V$), and $d \geq 0$ be an integer, then $\|f\|_{U^d(V)} \leq 1$, with equality if and only if f is of the form $f = e(P)$ for some polynomial $P : V \rightarrow \mathbb{T}$ of degree at most d , where $e : \mathbb{T} \rightarrow \mathbb{C}$ is the standard character $e(x) := e^{2\pi i x}$.

We collect the following standard facts about polynomials, setting V to equal the standard finite-dimensional vector space \mathbb{F}^n in order to use coordinates:

Lemma 1.7 (Basic facts about polynomials). *Let $V = \mathbb{F}^n$ for some natural number n , G be an additive group, and d be an integer.*

- (i) *If $P : V \rightarrow G$ is a function and $d \geq 0$, then $P \in \text{Poly}_{\leq d}(V \rightarrow G)$ if and only if $\Delta_h P \in \text{Poly}_{\leq d-1}(V \rightarrow G)$ for all $h \in V$. In fact, we may replace “for all $h \in V$ ” by “for all h in a set that generates V ”.*
- (ii) *If $d \geq 0$, then a function $P : V \rightarrow \mathbb{F}$ is a polynomial of degree $\leq d$ if and only if it has a representation of the form*

$$P(x_1, \dots, x_n) = \sum_{0 \leq i_1, \dots, i_n < p; i_1 + \dots + i_n \leq d} c_{i_1, \dots, i_n} x_1^{i_1} \dots x_n^{i_n} \quad (1.1)$$

for some coefficients $c_{i_1, \dots, i_n} \in \mathbb{F}$, and furthermore these coefficients are unique.

- (iii) *If $d \geq 0$, then a function $P : V \rightarrow \mathbb{T}$ is a polynomial of degree $\leq d$ if and only if it has a representation of the form*

$$P(x_1, \dots, x_n) = \alpha + \sum_{\substack{0 \leq i_1, \dots, i_n < p; j \geq 0 : \\ 0 < i_1 + \dots + i_n \leq d - j(p-1)}} \frac{c_{i_1, \dots, i_n, j} |x_1|^{i_1} \dots |x_n|^{i_n}}{p^{j+1}} \bmod 1 \quad (1.2)$$

for some coefficients $c_{i_1, \dots, i_n, j} \in \{0, 1, \dots, p-1\}$ and $\alpha \in \mathbb{T}$, where $x \mapsto |x|$ is the map from \mathbb{F} to the fundamental domain $\{0, 1, \dots, p-1\} \in \mathbb{Z}$. Furthermore the coefficients $c_{i_1, \dots, i_n, j}$ and α are unique.

- (iv) *If R is a commutative ring, and $P, Q : V \rightarrow R$ are polynomials of degree $\leq d$ and $\leq d'$ respectively, then $PQ : V \rightarrow R$ is a polynomial of degree $\leq d + d'$.*
- (v) *The map $p : P \mapsto pP$ is a homomorphism from*

$$\text{Poly}_{\leq d}(V \rightarrow G) \rightarrow \text{Poly}_{\leq \max(d-p+1, 0)}(V \rightarrow G).$$

If $G = \mathbb{T}$ and $d \geq 0$, then this homomorphism is surjective.

- (vi) *If $P \in \text{Poly}_{\leq d}(V \rightarrow \mathbb{T})$ and $d \geq 0$, then there exists $\alpha \in \mathbb{T}$ such that P takes values in the coset $\alpha + \frac{1}{p^{\lfloor \frac{d-1}{p-1} \rfloor + 1}} \mathbb{Z}/\mathbb{Z}$ of the $(p^{\lfloor \frac{d-1}{p-1} \rfloor + 1})^{\text{th}}$ roots of unity. In particular, P takes on at most $p^{\lfloor \frac{d-1}{p-1} \rfloor + 1}$ distinct values.*

Proof. See Appendix D. □

Remark 1.8. It follows from part (vi) that if $d < p$ then the set of polynomials of degree $\leq d$ in \mathbb{T} coincides (up to constants) with the set of classical polynomials of degree $\leq d$. However, this statement is false in the low characteristic case $p \leq d$.

Remark 1.9. We isolate one of the claims in the above lemma for special comment, namely the surjectivity claim in part (v). This claim implies that every polynomial

P of degree $\leq d$ for some $d \geq 0$ has a p^{th} root Q (i.e. $pQ = P$) which is a (non-classical) polynomial of degree $\leq d + p - 1$. We refer to the ability to take p^{th} roots while losing exactly a factor of $p - 1$ in the degree as the *exact roots property*. The exact roots property plays a crucial role in our proof of the inverse conjecture with the correct degree of polynomials involved. Unfortunately, this property does not hold in the ergodic theory setting, which is one reason why our arguments here do not proceed via the ergodic theoretic approach; see Appendix E for further discussion.

We now connect non-classical polynomials to the Gowers norms. From the monotonicity of the Gowers norms and the Cauchy-Schwarz-Gowers inequality (see Lemma B.1(ii), (vi)) we see that if $f : V \rightarrow \mathbb{C}$ correlates with a polynomial $P \in \text{Poly}_{\leq s+1}(V \rightarrow \mathbb{T})$ in the sense that

$$|\mathbb{E}_{x \in V} f(x) e(-P(x))| \geq \delta$$

for some $\delta > 0$, then we have $\|f\|_{U^{s+1}(V)} \geq \delta$.

The *inverse conjecture for the Gowers norm* over \mathbb{F} is a converse of this statement:

Conjecture 1.10 (Inverse conjecture $\text{GI}(s)$). *Let $\delta > 0$ and $s \geq 0$. Then there exists an $\epsilon = \epsilon_{\delta, s, \mathbb{F}} > 0$ such that for every finite-dimensional vector space V and any 1-bounded function $f : V \rightarrow \mathbb{C}$ with $\|f\|_{U^{s+1}(V)} \geq \delta$, there exists $P \in \text{Poly}_{\leq s}(V \rightarrow \mathbb{T})$ such that*

$$|\mathbb{E}_{x \in V} f(x) e(-P(x))| \geq \epsilon.$$

For a fixed s , we denote the above conjecture as $\text{GI}(s)$. We now briefly review the history of progress on this conjecture. The case $\text{GI}(0)$ is trivial, while the case $\text{GI}(1)$ follows easily from Plancherel's theorem. The result was established for $\text{GI}(2)$ in [18] (for odd characteristic) and [35] (for even characteristic), and a formulation of Conjecture 1.10 was then conjectured in both papers. In that formulation, the polynomial P was assumed to be a classical polynomial rather than a non-classical one. In subsequent work [22], [33], it was shown that this “classical” formulation of the conjecture could fail in the low characteristic regime $p \leq s$; however the counterexamples in these papers did not prevent the “non-classical” formulation of Conjecture 1.10 given above from holding in those cases.

The case when δ is sufficiently close to 1 (depending on s) was treated in [1], while the case when the characteristic p is large compared to s and δ was established in [36]. In [22], Conjecture 1.10 was also established in the case when $f = e(P)$ for some $P \in \text{Poly}_{< p}(V \rightarrow \mathbb{T})$. Finally, in [43], [3], Conjecture 1.10 was established in the high-characteristic case $s < p$, and a weaker version of this conjecture established in the low-characteristic case (see Theorem 11.3).

The first main result of this paper is to extend the high characteristic result from [43], [3] to the low characteristic case also:

Theorem 1.11. *$\text{GI}(s)$ is true for all choices of \mathbb{F} and s .*

Remark 1.12. As stated at the beginning of the introduction, we are restricting \mathbb{F} to be a field of prime order. But finite fields \mathbb{F}_{p^j} of prime power order are also covered

by this theorem, since any vector space over \mathbb{F}_{p^j} can also be viewed as a vector space over \mathbb{F}_p , with no change in the definition of the Gowers norm or the definition of a (non-classical) polynomial.

Remark 1.13. In [16], Gowers and Wolf used the high characteristic case of Conjecture 1.10 to compute the true complexity of a system (a concept introduced in [14]) of linear equations in sufficiently high characteristic. In principle, Theorem 1.11 would allow the “sufficiently high characteristic” condition to be weakened or dropped entirely. However, this would require adapting the arguments in [16] from classical polynomials to non-classical polynomials, and furthermore the high characteristic hypothesis is also used elsewhere in the arguments (in particular, the characteristic was assumed to exceed the “Cauchy-Schwarz complexity” of the system being studied).

1.14. Rank and analytic rank. Theorem 1.11 is established as a consequence of a related result (Theorem 1.20 below), which is the main technical result of the paper. Before we can state that result, we first need to recall the useful notions of rank and analytic rank, following [16].

Definition 1.15 (Rank and analytic rank). Let $s \geq 0$ be an integer, and let $P \in \text{Poly}_{\leq s+1}(V \rightarrow \mathbb{T})$.

- The *rank* $\text{rank}(P) = \text{rank}_s(P)$ of P is the least number m of polynomials $Q_1, \dots, Q_m \in \text{Poly}_{\leq s}(V \rightarrow \mathbb{T})$ of degree $\leq s$ such that P is a function of Q_1, \dots, Q_m , i.e. there exists a function $F : \mathbb{T}^m \rightarrow \mathbb{T}$ such that $P = F(Q_1, \dots, Q_m)$. (We adopt the convention that $\text{rank}_0(P)$ is infinite if the linear polynomial P is non-constant.)
- The *analytic rank* $\text{arank}(P) = \text{arank}_s(P)$ of P is defined to be the quantity $\text{arank}(P) := -\log_p \|e(P)\|_{U^{s+1}(V)}^{1/2^{s+1}}$.

We define the rank and analytic rank of polynomials $P \in \text{Poly}_{\leq s+1}(V \rightarrow \mathbb{F})$ by using the homomorphism ι_* , thus for instance $\text{arank}(P) := \text{arank}(\iota_* P)$. (In particular, when defining the rank of a classical polynomial P , we allow for P to be represented by *non-classical* polynomials Q_1, \dots, Q_m of the required degree.)

The analytic rank is closely related to the *derivative* $d^{s+1}P : V^{s+1} \rightarrow \mathbb{T}$ of P , defined as

$$d^{s+1}P(h_1, \dots, h_{s+1}) := \Delta_{h_1} \dots \Delta_{h_{s+1}} P(x) \quad (1.3)$$

for any $h_1, \dots, h_{s+1}, x \in V$ (note that the right-hand side is independent of x when P has degree at most $s+1$). Indeed, a short calculation shows that

$$\mathbb{E}_{h_1, \dots, h_{s+1} \in V} e(d^{s+1}P(h_1, \dots, h_{s+1})) = p^{-\text{arank}(P)}. \quad (1.4)$$

Thus for instance $\text{arank}(P)$ vanishes if and only if P is of degree $\leq s$.

For a classical quadratic form $Q : V \rightarrow \mathbb{F}$, the rank $\text{rank}(Q) = \text{rank}_1(Q)$ and the analytic rank $\text{arank}(Q) = \text{arank}_1(Q)$ are both equal (at least when p is odd) to the usual concept of the rank of a quadratic form in linear algebra; see [16] for further discussion.

Example 1.16. For each $k \geq 1$, let $S_k : \mathbb{F}_2^n \rightarrow \mathbb{F}_2$ be the degree k classical symmetric polynomial

$$S_k(x_1, \dots, x_n) := \sum_{1 \leq i_1 < \dots < i_k \leq n} x_{i_1} \cdots x_{i_k}$$

or equivalently, using the notation $L(x_1, \dots, x_n) := |x_1| + \dots + |x_n|$ from Example 1.6, we have

$$S_k = \binom{L}{k} \bmod 2.$$

A classical theorem of Lucas (reflecting the self-similar fractal nature of Pascal's triangle modulo 2) then gives the identity

$$S_k = S_{2^{a_1}} \cdots S_{2^{a_m}}$$

whenever $k = 2^{a_1} + \dots + 2^{a_m}$ is the binary expansion of k ; thus for instance $S_3 = S_2 S_1$, $S_5 = S_4 S_1$, $S_6 = S_4 S_2$, $S_7 = S_4 S_2 S_1$, etc. An easy induction on m then shows that each S^{2^m} is a binary coefficient of L , or more precisely that

$$L = \sum_{m=0}^{\infty} |S_{2^m}| 2^m \quad (1.5)$$

(note that only finitely many of the summands are non-zero).

In [33], [22] it was computed that

$$\|e(\iota(S_4))\|_{U^4(\mathbb{F}_2^n)}^4 = \frac{1}{8} + o(1)$$

as $n \rightarrow \infty$, and thus $\text{arank}(S_4) = \text{arank}_3(S_4) = 3 + o(1)$. As for the rank of S_4 , it can be shown that as $n \rightarrow \infty$, S_4 cannot be expressed as a function of a bounded number of *classical* cubics; see [33], [22]. However, from (1.5) we see that S_4 is a function of the expression $\frac{L}{8} \bmod 1$, which is a cubic by Lemma 1.7(iii). We conclude that $\text{rank}(S_4) = \text{rank}_3(S_4) = 1$.

We collect here some basic observations regarding rank and analytic rank:

Lemma 1.17 (Basic properties of rank). [16] *Let $s \geq 0$ be an integer, and let $P, Q \in \text{Poly}_{\leq s+1}(V \rightarrow \mathbb{T})$ be polynomials.*

- (i) $\text{rank}(P) = \text{rank}(-P)$ and $\text{arank}(P) = \text{arank}(-P)$.
- (ii) $\text{rank}(P + Q) \leq \text{rank}(P) + \text{rank}(Q)$ and $\text{arank}(P + Q) \leq 2^{s+1}(\text{arank}(P) + \text{arank}(Q))$.
- (iii) $\text{arank}(P) \leq C_s \text{rank}(P)$, for some constant $C_s > 0$ depending only on s .
- (iv) We have the inequality

$$|\mathbb{E}_{h_1, \dots, h_{s+1} \in V} e(d^{s+1} P(h_1, \dots, h_{s+1})) \prod_{j=1}^{s+1} f_j(h_1, \dots, h_{s+1})| \leq p^{-\text{arank}(P)/2^s}$$

whenever $f_j : V^{s+1} \rightarrow \mathbb{C}$ are 1-bounded functions, with each $f_j(h_1, \dots, h_{s+1})$ independent of the h_j variable.

Proof. The claim (i) is trivial, as is the first part of claim (ii). The second part of (ii) is established in [16, Lemma 5.9]. Now we turn to (iii). If we write $\text{rank}(P) = m$, then P is a function of polynomials Q_1, \dots, Q_m of degree $\leq s$. By Lemma 1.7(vi), we may assume that the Q_1, \dots, Q_m all take values in the $(p^C)^{\text{th}}$ roots of unity for some $C = C(s)$. By Fourier analysis, we may thus decompose

$$e(P) = \sum_{0 \leq a_1, \dots, a_m < p^C} c_{a_1, \dots, a_m} e(a_1 Q_1 + \dots + a_m Q_m)$$

where the Fourier coefficients c_{a_1, \dots, a_m} are bounded in magnitude by 1 (in fact their ℓ^2 norm is bounded by 1). By the pigeonhole principle, one can thus find a_1, \dots, a_m such that

$$|\mathbb{E}_{x \in V} e(P(x)) e(-a_1 Q_1(x) - \dots - a_m Q_m(x))| \geq p^{-mC}$$

which by the monotonicity of Gowers norms implies that

$$\|e(P) e(-a_1 Q_1 - \dots - a_m Q_m)\|_{U^{s+1}(V)} \geq p^{-mC}.$$

Since the $U^{s+1}(V)$ norm is invariant with respect to modulation by polynomials of degree $\leq s$, we conclude that

$$\|e(P)\|_{U^{s+1}(V)} \geq p^{-mC}.$$

and thus $\text{arank}(P) \leq 2^s C m$, and the claim follows.

Finally, the claim (iv) is established in [16, Lemma 5.4]. \square

Lemma 1.17(iii) asserts that the rank controls the analytic rank. For each $s \geq 0$, we let $\text{GIP}(s)$ denote the following converse:

Conjecture 1.18 (Inverse conjecture $\text{GIP}(s)$ for polynomials). *Let $\delta > 0$. Then there exists an integer $K = K_{\delta, s, \mathbb{F}} > 0$ such that for every finite-dimensional vector space V and any $Q \in \text{Poly}_{\leq s+1}(V \rightarrow \mathbb{C})$ with $\|e(Q)\|_{U^{s+1}(V)} \geq \delta$, there exists $P_1, \dots, P_K \in \text{Poly}_{\leq s}(V \rightarrow \mathbb{T})$ such that Q is a function of P_1, \dots, P_K (i.e. $Q = F(P_1, \dots, P_K)$ for some function $F : \mathbb{T}^K \rightarrow \mathbb{T}$).*

Remark 1.19. This type of conjecture was introduced by Bogdanov and Viola [5]. The case $\text{GIP}(0)$ is trivial, and the case $\text{GIP}(1)$ follows from Fourier analysis. In the high characteristic case $p > s + 1$, the result was established in [22, Proposition 6.1].

Our main technical theorem is then

Theorem 1.20. *$\text{GIP}(s)$ is true for all choices of \mathbb{F} and s .*

Theorem 1.11 can be deduced from Theorem 1.20 and the “weak” form of the inverse conjecture for the Gowers norm established in [43], [3], together with an argument from [16]; we give this (standard) argument in Section 11.

We remark that a different approach using ultrafilters to the structural theory of the Gowers norms is in the process of being carried out in [37], [38], [6].

We are indebted to the anonymous referee for a careful reading of the paper and many useful suggestions.

2. AN OUTLINE OF THE ARGUMENT

In this section we give an informal outline of how we will prove the inverse conjecture for polynomials $\text{GIP}(s)$ (and hence the full inverse conjecture $\text{GI}(s)$), suppressing many technical details (for instance, we will leave terms such as “bounded” vague for now, but such concepts will be made rigorous shortly with the assistance of nonstandard analysis). We will also identify \mathbb{F} with $\iota(\mathbb{F}) = \frac{1}{p}\mathbb{Z}/\mathbb{Z}$ for the purposes of this discussion.

As with many other arguments in this subject, we will induct on the degree parameter s , and assume that the conjecture $\text{GIP}(s')$ has already been proven for all $s' < s$. For sake of exposition we will work in the low characteristic case $p \leq s$, which is the hardest case.

Informally, the induction hypothesis allows one to obtain a completely satisfactory equidistribution theory for all (non-classical) polynomials of degree less than or equal to s , at least in principle. For instance, if $P : V \rightarrow \mathbb{T}$ is a polynomial of degree $s' + 1$ for some $s' < s$ that is of high rank, one can use $\text{GIP}(s')$ to describe the equidistribution of the tuples $(P(x + h_1\omega_1 + \dots + h_d\omega_d))_{\omega_1, \dots, \omega_d \in \{0,1\}^d} \in \mathbb{T}^{\{0,1\}^d}$ for fixed d , as x, h_1, \dots, h_d ranges uniformly over V . The precise equidistribution results we will need are rather technical to state, and will be formalised in Lemma 6.4 and Lemma 9.8.

The conjecture $\text{GIP}(s)$ asserts, informally, that any non-classical polynomial of degree $\leq s + 1$ of bounded analytic rank, also has bounded rank. This conjecture can be established from three sub-claims, which we informally state as follows:

- (i) (Multiplication by p) If P is a non-classical polynomial of degree $\leq s + 1$ of bounded analytic rank, then pP is a non-classical polynomial of degree $\leq s - p + 2$ of bounded analytic rank.
- (ii) (Division by p) If Q is a non-classical polynomial of degree $\leq s - p + 2$ of bounded rank, then there exists a non-classical polynomial P' of degree $\leq s + 1$ of bounded rank such that $pP' = Q$.
- (iii) (Classical case) If $P : V \rightarrow \mathbb{F}$ is a classical polynomial of degree $\leq s + 1$ of bounded analytic rank, then there exists a classical polynomial $Q : V \rightarrow \mathbb{F}$ of degree $\leq s + 1$ and bounded rank such that $d^{s+1}P = d^{s+1}Q$, where the top order derivative $d^{s+1}P : V^{s+1} \rightarrow \mathbb{F}$ of P was defined in (1.3). (In other words, P and Q differ by a polynomial of degree strictly less than $s + 1$.)

Indeed, assume these three claims hold. Then if P has bounded analytic rank, then by Claim (i), pP has bounded analytic rank, and hence is of bounded rank by induction hypothesis. By Claim (ii), we can thus find a polynomial P' of bounded rank (and therefore also of bounded analytic rank) such that $pP' = pP$, so that $P - P'$ is classical while still having bounded analytic rank. By Claim (iii), $P - P'$ differs from a classical

bounded rank polynomial Q by a polynomial of degree strictly less than $s + 1$, and so P has bounded rank also, as required.

It remains to verify the three claims. After performing a Fourier expansion, Claim (i) will be an easy consequence of the multidimensional Szemerédi theorem for vector spaces, first proven by Bergelson, Leibman, and McCutcheon [4]; we will establish it in Section 5.

We defer discussion of Claim (ii) for now, and move on to Claim (iii), which is somewhat easier to establish. The strategy here is to obtain as much structural information on the expression $d^{s+1}P$ as possible, so that the bounded rank polynomial Q can be constructed more or less explicitly. It is easy to see that $d^{s+1}P : V^{s+1} \rightarrow \mathbb{F}$ is a symmetric multilinear form. The fact that P is classical gives an additional cancellation property, namely that $d^{s+1}P(h_1, \dots, h_{s+1})$ necessarily vanishes whenever at least p of the h_1, \dots, h_{s+1} are equal to each other. For instance, if $p = 2$ and $s + 1 = 4$, we have

$$d^{s+1}P(a, a, b, c) = 0,$$

as can be deduced from the identity

$$\Delta_a \Delta_a P = (T_{2a} - 2T_a + 1)P = 2P - T_a(2P) = 0$$

since one has $2P = 0$ when P is classical. We will refer to symmetric multilinear forms with this cancellation property as *classical symmetric multilinear forms*.

The fact that P has bounded analytic rank implies (from (1.4)) that the form $d^{s+1}P$ is *biased*, in the sense that $\mathbb{E}_{h_1, \dots, h_{s+1} \in V} e(d^{s+1}P(h_1, \dots, h_{s+1}))$ is large. To exploit this, we use a general equidistribution result of Kaufman and Lovett [28] to conclude that $d^{s+1}P$ must be expressible in terms of lower degree classical symmetric multilinear forms. We then apply a standard “regularity lemma” (analogous to those in [22], [28]) to make these forms “independent” of each other, which makes them jointly equidistributed in a certain technical sense (see Lemma 7.8 for a precise statement). With this equidistribution, one can control the precise manner in which $d^{s+1}P$ is a function of the lower degree forms, and we will end up showing that $d^{s+1}P$ is a certain symmetrised combination of such forms (on a bounded index subspace). This will be made more precise in Section 6, but a typical example occurs when $s + 1 = 4$, in which the expression

$$d^4P(a, b, c, d)$$

will be expressed as a linear combination of terms such as

$$B(a, b)B(c, d) + B(a, c)B(b, d) + B(a, d)B(b, c) \tag{2.1}$$

where $B : V \times V \rightarrow \mathbb{F}$ is a (classical) symmetric bilinear form.

To conclude Claim (iii), we thus need to rewrite expressions such as (2.1) in the form $d^{s+1}Q$, where Q is of bounded rank. For sake of argument let us work specifically with the example (2.1). As B is itself classical, it can be expressed as $B = d^2R$ for some classical quadratic polynomial $R : V \rightarrow \mathbb{F}$. In the high characteristic case $p > 2$, one can then proceed simply by setting $Q := R^2/2!$, as the claim can be verified from the discrete Leibniz rule (D.2). However one cannot proceed so easily in the low characteristic case $p = 2$, as one can no longer divide by $2!$ in this case. Instead, we lift the polynomial

R (which takes values in $\mathbb{F} \equiv \mathbb{Z}/2\mathbb{Z}$) to the larger cyclic group $\mathbb{Z}/4\mathbb{Z}$, obtaining a cubic polynomial $\tilde{R} : V \rightarrow \mathbb{Z}/4\mathbb{Z}$ which projects back down to R in the sense that $\tilde{R} = R \bmod 2$. We then set $Q := \binom{\tilde{R}}{2} \bmod 2$ (instead of $R^2/2!$); the point is that the sequence $n \mapsto \binom{n}{2} \bmod 2$ is periodic with period 4 and so this expression is well defined. One can show that Q is a quartic polynomial, which is clearly of bounded rank as it depends only on the cubic polynomial \tilde{R} , and one can also compute that

$$d^4 Q(a, b, c, d) = B(a, b)B(c, d) + B(a, c)B(b, d) + B(a, d)B(b, c)$$

which gives the desired representation of (2.1). The same arguments work in more general degrees and characteristics to give Claim (iii) in general; see Section 6. Note that these constructions rely heavily on the classical nature of R , and hence of B and $d^4 P$, and ultimately exploit the multiplicative structure of the classical range \mathbb{F} that is not present in the non-classical range \mathbb{R}/\mathbb{Z} .

Finally, we return to Claim (ii). Let Q be a non-classical polynomial of degree $\leq s - p + 2$ of bounded rank, thus Q is some combination of polynomials of degree strictly less than $s - p + 2$. By a “regularity lemma” argument, it will turn out to be possible to express Q as the combination of “independent” polynomials P_1, \dots, P_k of various degrees between 1 and $s - p + 2$, with each P_j taking values in some cyclic group $\mathbb{Z}/p^{J_j}\mathbb{Z}$, thus

$$Q = F(P_1, \dots, P_k)$$

for some function $F : (\mathbb{Z}/p^{J_1}\mathbb{Z}) \times \dots \times (\mathbb{Z}/p^{J_k}\mathbb{Z}) \rightarrow \mathbb{T}$. The precise nature of this independence is somewhat technical to state, but it implies good joint equidistribution properties on the P_1, \dots, P_k (see Proposition 9.8 for a formal statement of this). The fact that Q has degree $\leq s - p + 2$ will imply that a suitable “weighted degree” of F is also at most $\leq s - p + 2$.

For technical reasons, the polynomials P_i of degree 1 will cause some difficulty (their p^{th} roots will not have high rank). But one can trivially eliminate all such polynomials by passing to a finite index subspace on which these polynomials are constant. As such, one can easily reduce to the case where all polynomials P_i have degree at least 2.

It would be convenient if we could then find another function $G : (\mathbb{Z}/p^{J_1}\mathbb{Z}) \times \dots \times (\mathbb{Z}/p^{J_k}\mathbb{Z}) \rightarrow \mathbb{T}$ of “weighted degree” $\leq s + 1$ such that $pG = F$, as the function $P' := G(P_1, \dots, P_k)$ would then obey the necessary requirements for Claim 2. Unfortunately, this claim turns out to be false in general. However, what one can do is first use the exact roots property from Lemma 1.7(v) to obtain a p^{th} root P'_j for each P_j with the “right” degree, thus $pP'_j = P_j$. Because the P_j have degree at least 2, it turns out that the independence properties of the P_j are inherited by the P'_j . One can then rewrite Q as $Q = F'(P'_1, \dots, P'_k)$ where $F' : (\mathbb{Z}/p^{J_1+1}\mathbb{Z}) \times \dots \times (\mathbb{Z}/p^{J_k+1}\mathbb{Z}) \rightarrow \mathbb{T}$ is the pullback of F . By analysing the concept of weighted degree for periodic functions on \mathbb{Z}^k (and in particular by breaking such functions down into multinomials), we will be able to find a function $G' : (\mathbb{Z}/p^{J_1+1}\mathbb{Z}) \times \dots \times (\mathbb{Z}/p^{J_k+1}\mathbb{Z}) \rightarrow \mathbb{T}$ of the right weighted degree such that $pG' = P$, and then the function $P' := G'(P'_1, \dots, P'_k)$ will obey the properties required for Claim (ii).

The detailed proof of Claim (ii) will occupy Sections 9-10.

3. TAKING ULTRALIMITS

To prove Theorem 1.20 it will be convenient to pass from the finitary setting to an infinitary one, in order to eliminate a profusion of epsilons, deltas, and growth functions (such as the growth functions \mathcal{F} that appear for instance in [22], [28]); it also allows us to conveniently make rigorous such phrases as “the polynomials P_1, \dots, P_m are linearly independent modulo bounded rank errors”, which would otherwise only make sense heuristically or would need to be quantified with additional parameters. In [43] the Furstenberg correspondence principle was used to convert the inverse conjecture to a statement in an infinitary branch of mathematics, namely ergodic theory. Unfortunately, the ergodic theory framework has a drawback in the low characteristic setting, namely that there does not appear to be an easy way to take roots of polynomials in this setting in a degree-efficient manner (see Appendix E). To overcome this technical obstacle we shall take a different infinitary formulation of the problem, namely an *ultralimit* (or *nonstandard analysis*) formulation. (Such formulations have also appeared in other recent work on the inverse conjecture [24], [37].)

The basic machinery of ultralimits and nonstandard analysis is recalled in Appendix A, as is the asymptotic notation (such as $X \ll Y$ or $X = O(Y)$) associated with this machinery.

We now translate all of the terminology used for Theorem 1.20 to the ultralimit setting.

Let V be a non-empty limit finite set (i.e. an ultralimit $V = \lim_{\mathbf{n} \rightarrow \alpha} V_{\mathbf{n}}$ of standard non-empty finite sets $V_{\mathbf{n}}$) and let $f : V \rightarrow {}^*\mathbb{C}$ be a limit function on V (thus f is an ultralimit $f = \lim_{\mathbf{n} \rightarrow \alpha} f_{\mathbf{n}}$ of standard functions $f_{\mathbf{n}} : V_{\mathbf{n}} \rightarrow \mathbb{C}$). Then we can define the expectation $\mathbb{E}_{x \in V} f(x)$ of f on V in the usual fashion by the formula

$$\mathbb{E}_{x \in V} f(x) := \lim_{\mathbf{n} \rightarrow \alpha} \mathbb{E}_{x_{\mathbf{n}} \in V_{\mathbf{n}}} f_{\mathbf{n}}(x_{\mathbf{n}}).$$

This will be a limit complex number. If further V is a limit finite-dimensional vector space (i.e. each $V_{\mathbf{n}}$ is a standard finite-dimensional space), and $d \geq 1$ is a standard natural number, then we can similarly define the uniformity norm

$$\|f\|_{U^d(V)} := \lim_{\mathbf{n} \rightarrow \alpha} \|f_{\mathbf{n}}\|_{U^d(V_{\mathbf{n}})}$$

which will be a limit non-negative real number.

Given a limit finite-dimensional space V and a standard integer $d \in \mathbb{Z}$, we let $\text{Poly}_{\leq d}(V \rightarrow {}^*\mathbb{T})$ denote the space of all limit polynomials P of degree $\leq d$ from V to ${}^*\mathbb{T}$, i.e. all ultralimits $P = \lim_{\mathbf{n} \rightarrow \alpha} P_{\mathbf{n}}$ of polynomials $P_{\mathbf{n}} \in \text{Poly}_{\leq d}(V_{\mathbf{n}} \rightarrow \mathbb{T})$. We define $\text{Poly}_{\leq d}(V \rightarrow \mathbb{F})$ and $\text{Poly}_{\leq d}(V \rightarrow \iota(\mathbb{F}))$ for limit finite-dimensional V similarly.

We make the simple but important remark that all the claims in Lemma 1.7, that were established for finite-dimensional vector spaces, extend to the limit finite-dimensional setting (replacing all operations by their limit counterparts) by taking ultralimits. In particular, the surjectivity claim in Lemma 1.7(v) extends to the limit finite-dimensional setting.

If $s \geq 0$ is standard and $P \in \text{Poly}_{\leq s+1}(V \rightarrow {}^*\mathbb{T})$, we can define the rank, $\text{rank}(P) = \text{rank}_s(P)$, and analytic rank, $\text{arank}(P) = \text{arank}_s(P)$, which are now non-negative limit integers and non-negative limit real numbers respectively. We let $\text{Poly}_{\leq s+1, \text{BR}}(V \rightarrow {}^*\mathbb{T})$ and $\text{Poly}_{\leq s+1, \text{BAR}}(V \rightarrow {}^*\mathbb{T})$ denote the limit polynomials $P \in \text{Poly}_{\leq s+1}(V \rightarrow {}^*\mathbb{T})$ of degree $\leq s+1$ that are of bounded rank and bounded analytic rank respectively. From Lemma 1.17 we see that $\text{Poly}_{\leq s+1, \text{BR}}(V \rightarrow {}^*\mathbb{T})$ and $\text{Poly}_{\leq s+1, \text{BAR}}(V \rightarrow {}^*\mathbb{T})$ are vector spaces with the inclusions

$$\text{Poly}_{\leq s}(V \rightarrow {}^*\mathbb{T}) \leq \text{Poly}_{\leq s+1, \text{BR}}(V \rightarrow {}^*\mathbb{T}) \leq \text{Poly}_{\leq s+1, \text{BAR}}(V \rightarrow {}^*\mathbb{T}) \leq \text{Poly}_{\leq s+1}(V \rightarrow {}^*\mathbb{T}).$$

We similarly define $\text{Poly}_{\leq s+1, \text{BR}}(V \rightarrow \mathbb{F})$ and $\text{Poly}_{\leq s+1, \text{BAR}}(V \rightarrow \mathbb{F})$.

We can now give the ultralimit formulation of GIP(s):

Theorem 3.1 (Ultralimit equivalence of GIP(s)). *Let $s \geq 0$ be standard. Then GIP(s) holds if and only if, for every limit finite-dimensional vector space V , $\text{Poly}_{\leq s+1, \text{BR}}(V \rightarrow {}^*\mathbb{T}) = \text{Poly}_{\leq s+1, \text{BAR}}(V \rightarrow {}^*\mathbb{T})$ (i.e. bounded analytic rank and bounded rank are equivalent).*

Proof. We first assume GIP(s) and verify that $\text{Poly}_{\leq s+1, \text{BR}}(V \rightarrow {}^*\mathbb{T}) = \text{Poly}_{\leq s+1, \text{BAR}}(V \rightarrow {}^*\mathbb{T})$. Accordingly, we let $P \in \text{Poly}_{\leq s+1, \text{BAR}}(V \rightarrow {}^*\mathbb{T})$ and need to show that P has bounded rank.

By construction, we can write $V = \prod_{\mathbf{n} \rightarrow \alpha} V_{\mathbf{n}}$ as the ultraproduct of finite-dimensional spaces $V_{\mathbf{n}}$, and similarly write $P = \lim_{\mathbf{n} \rightarrow \alpha} P_{\mathbf{n}}$ as the ultralimit of polynomials $P_{\mathbf{n}} : V_{\mathbf{n}} \rightarrow \mathbb{T}$. Since P has bounded analytic rank, the $P_{\mathbf{n}}$ have bounded analytic rank uniformly in \mathbf{n} (at least for \mathbf{n} sufficiently close to α). Applying GIP(s), we conclude that the $P_{\mathbf{n}}$ have bounded rank uniformly in \mathbf{n} , thus we can find polynomials $Q_{\mathbf{n},1}, \dots, Q_{\mathbf{n},K} : V_{\mathbf{n}} \rightarrow \mathbb{T}$ of degree $\leq s$ and a function $F_{\mathbf{n}} : \mathbb{T}^K \rightarrow \mathbb{T}$ such that $P_{\mathbf{n}} = F_{\mathbf{n}}(Q_{\mathbf{n},1}, \dots, Q_{\mathbf{n},K})$. Writing $Q_i := \lim_{\mathbf{n} \rightarrow \alpha} Q_{\mathbf{n},i}$ and $F := \lim_{\mathbf{n} \rightarrow \alpha} F_{\mathbf{n}}$, we conclude that $Q_1, \dots, Q_K : V \rightarrow {}^*\mathbb{T}$ are limit polynomials of degree $\leq s$ and $P = F(Q_1, \dots, Q_K)$, and so P has bounded rank as desired.

Conversely, suppose that $\text{Poly}_{\leq s+1, \text{BR}}(V \rightarrow {}^*\mathbb{T}) = \text{Poly}_{\leq s+1, \text{BAR}}(V \rightarrow {}^*\mathbb{T})$. We assume for contradiction that GIP(s) failed. Thus, there exists a sequence $V_{\mathbf{n}}$ of finite-dimensional vector spaces and polynomials $P_{\mathbf{n}} : V_{\mathbf{n}} \rightarrow \mathbb{T}$ of degree $\leq s+1$ whose analytic rank is bounded uniformly in \mathbf{n} , but whose rank goes to infinity as $\mathbf{n} \rightarrow \infty$. Setting $V := \prod_{\mathbf{n} \rightarrow \alpha} V_{\mathbf{n}}$ and $P := \lim_{\mathbf{n} \rightarrow \alpha} P_{\mathbf{n}}$, we see that $P : V \rightarrow {}^*\mathbb{T}$ is a limit polynomial of degree $\leq s+1$ of bounded analytic rank, and hence of bounded rank by hypothesis, thus $P = F(Q_1, \dots, Q_K)$ for some bounded m , some limit polynomials $Q_1, \dots, Q_K : V \rightarrow {}^*\mathbb{T}$ of degree $\leq s$, and some function $F : {}^*\mathbb{T}^K \rightarrow {}^*\mathbb{T}$. Writing $Q_i := \lim_{\mathbf{n} \rightarrow \alpha} Q_{\mathbf{n},i}$ and $F := \lim_{\mathbf{n} \rightarrow \alpha} F_{\mathbf{n}}$, we conclude that $P_{\mathbf{n}} = F_{\mathbf{n}}(Q_{\mathbf{n},1}, \dots, Q_{\mathbf{n},m})$ for all \mathbf{n} sufficiently close to α , and thus the rank of $P_{\mathbf{n}}$ is bounded uniformly for such \mathbf{n} , which gives the desired contradiction. \square

A similar argument (which we omit) gives the ultralimit formulation of GI(s):

Theorem 3.2. *Let $s \geq 0$ be standard. Then GI(s) holds if and only if, for every limit finite-dimensional vector space V and every bounded limit function $f : V \rightarrow {}^*\mathbb{C}$ (where*

“bounded” means that $\sup_{x \in V} |f(x)|$ is bounded), $\|f\|_{U^{s+1}(V)} \gg 1$ if and only if there exists $P \in \text{Poly}_{\leq s}(V \rightarrow {}^*\mathbb{T})$ such that $|\mathbb{E}_{x \in V} f(x)e(-P(x))| \gg 1$.

It remains to establish that $\text{Poly}_{\leq s+1, \text{BR}}(V \rightarrow {}^*\mathbb{T}) = \text{Poly}_{\leq s+1, \text{BAR}}(V \rightarrow {}^*\mathbb{T})$. This is the purpose of the remaining sections of the paper.

4. SPLITTING INTO THREE SUBCLAIMS

To prove the claim $\text{Poly}_{\leq s+1, \text{BR}}(V \rightarrow {}^*\mathbb{T}) = \text{Poly}_{\leq s+1, \text{BAR}}(V \rightarrow {}^*\mathbb{T})$ (and hence $\text{GIP}(s)$), we split this theorem into three subclaims as outlined in Section 2. Claim (i) is easily formalised:

Theorem 4.1 (Multiplication by p). *Let V be a limit finite-dimensional vector space. If $k > p$ is a standard integer, then the map $P \mapsto pP$ maps $\text{Poly}_{\leq k, \text{BAR}}(V \rightarrow {}^*\mathbb{T})$ to $\text{Poly}_{\leq k-p+1, \text{BAR}}(V \rightarrow {}^*\mathbb{T})$.*

We prove this theorem in Section 5. Claim (ii) is also easily formalised, though for technical reasons (having to do with the need to eliminate all linear polynomials that arise in the regularity lemma) it is convenient to weaken the claim to a “virtual” version in which one only obtains roots on a bounded index subspace:

Theorem 4.2 (Exact roots). *Let $s \geq 1$ be a standard integer such that $\text{GIP}(s')$ is true for all $0 \leq s' \leq s$. Then for every $0 \leq s' \leq s$, every limit finite-dimensional V , and every $P \in \text{Poly}_{\leq s'+1, \text{BR}}(V \rightarrow {}^*\mathbb{T})$, there exists a bounded index (limit) subspace V' of V and $Q \in \text{Poly}_{\leq s'+p, \text{BR}}(V' \rightarrow {}^*\mathbb{T})$ such that $pQ = P$ on V' .*

We will prove this theorem in Sections 9-10.

To formalise Claim (iii) properly, we will need some additional notation.

Definition 4.3 (Multilinear maps). Let V be a limit finite-dimensional vector space and $k \geq 1$.

- A k -linear map is a limit map $T : V^k \rightarrow \mathbb{F}$ such that if one fixes all but one variable h_i of the k variables $h_1, \dots, h_k \in V$, the remaining map $h_i \mapsto T(h_1, \dots, h_k)$ is linear.
- A k -linear map is *symmetric* if it is invariant under permutations of the k variables h_1, \dots, h_k .
- A symmetric k -linear map is *classical* if $T(h_1, \dots, h_k)$ vanishes whenever at least p of the h_1, \dots, h_k agree (this condition is of course vacuous for $k < p$). Thus, for instance, if $p = 2$ and $k = 3$, a symmetric trilinear form T is classical if $T(a, a, b) = 0$ for all $a, b \in V$.

We abbreviate “classical symmetric multilinear” as CSM. We denote the space of classical symmetric k -linear maps $T : V^k \rightarrow \mathbb{F}$ as $\text{CSM}^k(V)$; this is clearly a vector space.

Example 4.4. Let $V = \mathbb{F}^n$ and $k = 2$. The map

$$T(h, h') := \sum_{1 \leq i < j \leq n} h_i h'_j$$

where $h = (h_1, \dots, h_n), h' = (h'_1, \dots, h'_n)$ is multilinear, but not symmetric or classical. The map

$$T'(h, h') := \sum_{1 \leq i, j \leq n} h_i h'_j$$

is symmetric and multilinear, but not classical. The map

$$T''(h, h') := \sum_{1 \leq i, j \leq n: i \neq j} h_i h'_j$$

is symmetric, multilinear, and classical, and thus lies in $\text{CSM}^2(V)$.

Multilinear maps are naturally associated to derivatives of polynomials. Indeed, from the cocycle equation

$$\Delta_{h+h'} = \Delta_h + \Delta_{h'} + \Delta_h \Delta_{h'}$$

and the commutativity identity

$$\Delta_h \Delta_{h'} = \Delta_{h'} \Delta_h$$

we see that for any $P \in \text{Poly}_{\leq k}(V \rightarrow {}^*\mathbb{T})$, $d^k P$ is multilinear and symmetric. In particular, from multilinearity $d^k P$ must take values in $\iota(\mathbb{F})$.

From Lemma D.3 we see that Δ_h^p and $p\Delta_h$ differ (multiplicatively) by an invertible formal differential operator, which is equal to -1 plus higher order terms. Applying this to a polynomial $P \in \text{Poly}_{\leq k}(V \rightarrow {}^*\mathbb{T})$ with $k \geq p$, we conclude the identity

$$d^k P(h_1, \dots, h_1, h_2, \dots, h_{k-p+1}) = -d^{k-p+1}(pP)(h_1, h_2, \dots, h_{k-p+1}) \quad (4.1)$$

for any $h_1, \dots, h_{k-p+1} \in V$, where h_1 appears p times on the left-hand side (recall that pP is of degree $\leq k - p + 1$, by Lemma 1.7).

This identity has a number of consequences. For instance, it gives some additional constraints on $d^k P$ beyond symmetry, such as

$$d^k P(h_1, \dots, h_1, h_2, \dots, h_{k-p+1}) = d^k P(h_1, h_2, \dots, h_2, h_3, \dots, h_{k-p+1}),$$

where p copies of h_1 appear on the left, and p copies of h_2 appear on the right.

Another consequence will be important for us:

Lemma 4.5 (Derivative of classical polynomials). *Let $k \geq 1$ and V be a limit finite-dimensional vector space, then d^k maps $\text{Poly}_{\leq k}(V \rightarrow \mathbb{F})$ to $\text{CSM}^k(V \rightarrow \mathbb{F})$, and furthermore this map is surjective. In other words, we have the short exact sequence*

$$0 \rightarrow \text{Poly}_{\leq k-1}(V \rightarrow \mathbb{F}) \rightarrow \text{Poly}_{\leq k}(V \rightarrow \mathbb{F}) \xrightarrow{d^k} \text{CSM}^k(V) \rightarrow 0.$$

To state the above lemma loosely, classical symmetric multilinear forms are nothing more than the derivatives of classical polynomials.

Proof. Applying (4.1) to a classical polynomial $P \in \text{Poly}_{\leq k}(V \rightarrow \mathbb{F})$, we have $pP = 0$, and thus $d^k P$ is classical, which gives the first claim.

Conversely, suppose that $T \in \text{CSM}^k(V)$. It suffices to verify the claim when V is finite dimensional, as the limit finite-dimensional case then follows by taking ultralimits. We take advantage of the finite-dimensionality to write $V = \mathbb{F}^n$ (without loss of generality), then the multilinear form T can be expressed in coordinates as

$$T(h_1, \dots, h_k) = \iota \left(\sum_{1 \leq i_1, \dots, i_k \leq n} c_{i_1, \dots, i_k} h_{1, i_1} \dots h_{k, i_k} \right)$$

where $h_{i,1}, \dots, h_{i,n} \in \mathbb{F}$ are the coordinates of h_i , and $c_{i_1, \dots, i_k} \in \mathbb{F}$ are coefficients. From the symmetric nature of T we know that the c_{i_1, \dots, i_k} are symmetric with respect to permutations of the indices i_1, \dots, i_k ; from the classical nature of T we know that the c_{i_1, \dots, i_k} vanish whenever p or more of the i_j are equal. We thus see that T is an integer linear combination of expressions of the form

$$\iota \left(\sum_{\{i_1, \dots, i_k\} = A} h_{1, i_1} \dots h_{k, i_k} \right)$$

where $\{i_1, \dots, i_k\}$ is the multiset formed by i_1, \dots, i_k , and A is a multiset of k elements taking values in $\{1, \dots, n\}$, with the multiplicity a_j of each $1 \leq j \leq n$ in A being less than p . A short computation then shows that each such expression can be expressed as $d^k P$ for a polynomial $P \in \text{Poly}_{\leq k}(V \rightarrow \mathbb{F})$, indeed we may take

$$P(x_1, \dots, x_n) := \iota \left(\prod_{j=1}^n \frac{x_j^{a_j}}{a_j!} \right).$$

Note how the multiplicity bound $a_j < p$ allows for the factorial $a_j!$ to be inverted in \mathbb{F} . (One could also use the binomial coefficient $\binom{x_j}{a_j}$ in place of $\frac{x_j^{a_j}}{a_j!}$ here if desired.) The claim now follows from linearity of d^k . \square

Claim (iii) will be a variant of Lemma 4.5 in the biased case:

Theorem 4.6 (Inverse Gowers for classical symmetric multilinear forms). *Let $s \geq 2$ be such that $\text{GIP}(s')$ is true for all $0 \leq s' < s$, let V be limit finite-dimensional, and let $T \in \text{CSM}^{s+1}(V \rightarrow \mathbb{F})$ be such that $|\mathbb{E}_{h_1, \dots, h_{s+1} \in V}(\iota(T(h_1, \dots, h_{s+1})))| \gg 1$. Then there exists a bounded index (limit) subspace V' of V and $P \in \text{Poly}_{\leq s+1, \text{BR}}(V' \rightarrow \mathbb{F})$ such that $d^{s+1}P = T$ on $(V')^{s+1}$.*

This theorem will be proven in Sections 6-8. To state the theorem loosely, classical *biased* symmetric multilinear forms are nothing more than the derivatives of classical *bounded rank* polynomials.

In the remainder of this section, we show how Theorems 4.1, 4.2, 4.6 imply $\text{GIP}(s)$. We first need a technical lemma to handle the passage to bounded index limit subspaces.

Lemma 4.7. *Let $k \geq 2$ and $P \in \text{Poly}_{\leq k}(V \rightarrow {}^*\mathbb{T})$, and let V' be a bounded index subspace of V . (Note that such subspaces are automatically limit subspaces.) Let $P' \in \text{Poly}_{\leq k}(V' \rightarrow {}^*\mathbb{T})$ be the restriction of P to V' .*

- (i) *P has bounded analytic rank if and only if P' has bounded analytic rank.*
- (ii) *P has bounded rank if and only if P' has bounded rank.*

If $k = 1$, then the “only if” portions of the claim continue to hold.

Proof. We begin with (i). If P' has bounded analytic rank, then by (1.4)

$$|\mathbb{E}_{h_1, \dots, h_k \in V} 1_{V'}(h_1) \dots 1_{V'}(h_k) e(d^k P(h_1, \dots, h_k))| \gg 1.$$

Applying Lemma 1.17(iv) we conclude that P has bounded analytic rank, as required.

Conversely, suppose that P has bounded analytic rank, then by (1.4) we have

$$\mathbb{E}_{h_1, \dots, h_k \in V} e(d^k P(h_1, \dots, h_k)) \gg 1.$$

Applying Fourier analysis, we conclude that for $\gg |V|^{k-1}$ $k-1$ -tuples (h_1, \dots, h_{k-1}) in V^{k-1} , one has $d^k P(h_1, \dots, h_{k-1}, \cdot) \equiv 0$. Applying the pigeonhole principle, one can restrict h_{k-1} to a single coset of V' and still have the above claim. Using multilinearity in h_{k-1} , one can in fact restrict h_{k-1} to V' . Iterating this argument one may restrict all of h_1, \dots, h_{k-1} to V' . Using Fourier analysis we conclude that

$$\mathbb{E}_{h_1, \dots, h_k \in V'} e(d^k P'(h_1, \dots, h_k)) \gg 1,$$

and hence by (1.4), P has bounded analytic rank as desired.

Now we show (ii). The “only if” portion is trivial, so we focus on the “if” part. By induction we may assume that V' is a hyperplane in V , and we may then write $V = V' \times \mathbb{F}$ without loss of generality. Suppose that P' is of bounded rank. Letting $\pi : V \rightarrow V'$ be the coordinate projection from $V' \times \mathbb{F}$ to V' , it is easy to verify that $P' \circ \pi$ is a bounded rank polynomial of degree $\leq k$. Subtracting this from P , we may assume without loss of generality that P vanishes on V' .

On each coset of $h + V'$ of V' , P is of degree $\leq k-1$ (since $\Delta_h P$ is of degree $\leq k-1$). Thus $1_{h+V'} P$ is a function of finitely many polynomials of degree $\leq k-1$. Summing over a set of coset representatives h we obtain the claim. \square

Now we can prove $\text{GIP}(s)$ assuming Theorems 4.1, 4.2, 4.6.

Proof of $\text{GIP}(s)$. By induction we may assume that $\text{GIP}(s')$ holds for all $0 \leq s' < s$; our task is now to show that $\text{GIP}(s)$ holds. We will assume $s \geq 2$, as the $s = 0, 1$ cases are well known.

Let $P \in \text{Poly}_{\leq s+1, \text{BAR}}(V \rightarrow {}^*\mathbb{T})$. By Theorem 3.1, it will suffice to show that P has bounded rank.

Let us first suppose that $s + 1 > p$ (which is the most difficult case). By Theorem 4.1, $pP \in \text{Poly}_{\leq s-p+2, \text{BAR}}(V \rightarrow {}^*\mathbb{T})$, so by GIP($s - p + 1$), pP is of bounded rank. By Theorem 4.2 (with s replaced by $s - 1$), there thus exists $Q \in \text{Poly}_{\leq s+1, \text{BR}}(V' \rightarrow {}^*\mathbb{T}) \subset \text{Poly}_{\leq s+1, \text{BAR}}(V' \rightarrow {}^*\mathbb{T})$ such that $pP = pQ$ on a bounded index subspace V' of V . In particular, $P - Q$ takes values in $\iota(\mathbb{F})$ on V' , so by Lemma 4.5, $d^{s+1}(P - Q) \in \text{CSM}^{s+1}(V' \rightarrow {}^*\mathbb{T})$.

By Lemma 4.7 P, Q both lie in $\text{Poly}_{\leq s+1, \text{BAR}}(V \rightarrow {}^*\mathbb{T})$, and so $P - Q$ does also. Thus

$$\mathbb{E}_{h_1, \dots, h_{s+1} \in V} e(d^{s+1}(P - Q)(h_1, \dots, h_{s+1})) \gg 1.$$

Applying Theorem 4.6 (pulling back by ι) we conclude that $d^{s+1}(P - Q) = d^{s+1}(W)$ on $(V'')^{s+1}$ for some $W \in \text{Poly}_{\leq s+1, \text{BR}}(V'' \rightarrow \iota(\mathbb{F}))$ on some bounded index subspace V'' of V , thus $P - Q - W$ is a polynomial of degree $\leq s + 1$ on $V' \cap V''$. Since Q, W are of bounded rank on $V' \cap V''$, we conclude that P is of bounded rank on $V' \cap V''$, and hence on V by Lemma 4.7.

The case $s < p$ is similar. Here, $pP \in \text{Poly}_{\leq 1}(V \rightarrow {}^*\mathbb{T})$, hence on passing to a subspace pP is constant, and in particular we can find a constant Q such that $pP = pQ$ on this subspace. One then argues as before. \square

It remains to prove Theorems 4.1, 4.2, 4.6. This is the purpose of the remaining sections of the paper.

5. MULTIPLYING BY p

In this section we prove Theorem 4.1. The main tool will be the multidimensional Szemerédi theorem over finite fields² from [4], which we can formulate as follows:

Proposition 5.1 (Multidimensional Szemerédi theorem). *Let V be a limit finite-dimensional vector space, let $k \geq 1$ be a standard integer, and let $A \subset V^k$ be such that $|A| \gg |V|^k$. Then there exist $\gg |V|^{k+1}$ tuples $(h_1, \dots, h_k, h) \in V$ such that*

$$\{(h_1 + h\omega_1, h_2 + h\omega_2, \dots, h_k + h\omega_k) : \omega_1, \dots, \omega_k \in \mathbb{F}\} \subset A.$$

Proof. We combine the results from [4] with an averaging argument of Varnavides [45].

The claim is trivial if V has bounded dimension, so we may assume that V has unbounded dimension.

Since $|A| \gg |V|^k$, if we write $A = \lim_{\mathbf{n} \rightarrow \alpha} A_{\mathbf{n}}$ and $V = \lim_{\mathbf{n} \rightarrow \alpha} V_{\mathbf{n}}$, then there is a standard $\delta > 0$ such that $|A_{\mathbf{n}}| \geq \delta |V_{\mathbf{n}}|^k$ for all \mathbf{n} sufficiently close to α .

Fix \mathbf{n} with this property. Let M be a large standard integer depending on k, δ to be chosen later. As V has unbounded dimension, we may assume that $V_{\mathbf{n}}$ has dimension at least M by taking \mathbf{n} sufficiently close to α .

²It is also possible to prove this proposition using the density Hales-Jewett theorem[9] instead.

Let $W_{\mathbf{n}}$ be an arbitrary linear subspace of $V_{\mathbf{n}}$ of dimension M . We can then foliate $V_{\mathbf{n}}^k$ into $(|V_{\mathbf{n}}|/|W_{\mathbf{n}}|)^k$ cosets $x + W_{\mathbf{n}}^k$ of $W_{\mathbf{n}}^k$. The average value of $|A_{\mathbf{n}} \cap (x + W_{\mathbf{n}}^k)|/|W_{\mathbf{n}}|^k$ in these cosets is at least δ . Thus, for at least $\delta(|V_{\mathbf{n}}|/|W_{\mathbf{n}}|)^k/2$ of these cosets, we have $|A_{\mathbf{n}} \cap (x + W_{\mathbf{n}}^k)|/|W_{\mathbf{n}}|^k \geq \delta/2$.

Consider one of these cosets $x + W_{\mathbf{n}}^k$ with $|A_{\mathbf{n}} \cap (x + W_{\mathbf{n}}^k)|/|W_{\mathbf{n}}|^k \geq \delta/2$. Applying [4, Corollary 5.4], we see that if M is large enough depending on k, δ , we can thus find a non-zero $h \in W_{\mathbf{n}}$ and $(h_1, \dots, h_k) \in x + W_{\mathbf{n}}^k$ such that

$$\{(h_1 + h\omega_1, h_2 + h\omega_2, \dots, h_k + h\omega_k) : \omega_1, \dots, \omega_k \in \mathbb{F}\} \subset A_{\mathbf{n}}. \quad (5.1)$$

Summing over all such cosets (and using the pigeonhole principle to fix h), we can thus find a non-zero $h \in W_{\mathbf{n}}$ such that (5.1) holds for at least $c_{k,M,\delta}|V_{\mathbf{n}}|^k$ tuples $(h_1, \dots, h_k) \in V_{\mathbf{n}}^k$, where $c_{k,M,\delta} > 0$ is standard. Averaging over all possible linear subspaces $W_{\mathbf{n}}$ of $V_{\mathbf{n}}$ of dimension M , we conclude from a routine double counting argument that in fact (5.1) holds for at least $c'_{k,M,\delta}|V_{\mathbf{n}}|^{k+1}$ tuples $(h, h_1, \dots, h_k) \in V_{\mathbf{n}}^{k+1}$, where $c'_{k,M,\delta} > 0$ is standard. Taking ultralimits as $\mathbf{n} \rightarrow \alpha$ we obtain the claim. \square

Now we can prove Theorem 4.1.

Proof of Theorem 4.1. Let $P \in \text{Poly}_{\leq k, \text{BAR}}(V \rightarrow {}^*\mathbb{T})$. From (1.4) one has

$$\mathbb{E}_{h_1, \dots, h_k \in V} e(d^k P(h_1, \dots, h_k)) \gg 1. \quad (5.2)$$

For each tuple $(h_1, \dots, h_{k-1}) \in V^{k-1}$, the function $e(d^k P(h_1, \dots, h_{k-1}, \cdot))$ is a character on V . Thus, by Fourier analysis, for (5.2) to hold we must have $\gg |V|^{k-1}$ tuples $(h_1, \dots, h_{k-1}) \in V^{k-1}$, for which the form

$$d^k P(h_1, \dots, h_{k-1}, \cdot)$$

vanishes identically. Applying Proposition 5.1, we can thus find $\gg |V|^k$ tuples $(h, h_1, \dots, h_{k-1}) \in V^k$ such that

$$d^k P(h_1 + \omega_1 h, \dots, h_p + \omega_p h, h_{p+1}, \dots, h_{k-1}, \cdot)$$

vanishes identically for all $(\omega_1, \dots, \omega_p) \in \{0, 1\}^p$. (Note how the hypothesis $k > p$ was needed here in order for this expression to make sense.) Taking an alternating sum of these expressions, we thus have

$$d^k P(h, \dots, h, h_{p+1}, \dots, h_{k-1}, \cdot)$$

vanishing identically for $\gg |V|^k$ tuples $(h, h_1, \dots, h_{k-1}) \in V^k$, and thus for $\gg |V|^{k-p+1}$ tuples $(h, h_p, \dots, h_{k-1}) \in V^{k-p+1}$. Once again this implies that

$$\mathbb{E}_{h, h_{p+1}, \dots, h_k \in V} e(d^k P(h, \dots, h, h_{p+1}, \dots, h_{k-1}, h_k)) \gg 1,$$

and the claim follows from (1.4), and (4.1). \square

For future use, we record an immediate corollary of Theorem 4.1:

Corollary 5.2. *Suppose that $k > p$ and $\text{GIP}(k-p)$ holds. Then the map $p : P \mapsto pP$ maps $\text{Poly}_{\leq k, \text{BR}}(V \rightarrow {}^*\mathbb{T})$ to $\text{Poly}_{\leq k-p+1, \text{BR}}(V \rightarrow {}^*\mathbb{T})$.*

We will use Corollary 5.2 in the following contrapositive sense: if $k > p$, $\text{GIP}(k - p)$ holds, and $P \in \text{Poly}_{\leq k}(V \rightarrow {}^*\mathbb{T})$ is such that pP has unbounded rank (as a polynomial of degree $\leq k - p + 1$), then P has unbounded rank (as a polynomial of degree $\leq k$). Informally, roots of high rank nonlinear polynomials remain high rank.

We remark that Theorem 4.1 (and Corollary 5.2) fail when $k = p$. For instance, if $p = 2$ and $V = \mathbb{F}_2$, then the function $P(x) := |x|/4 \bmod 1$ is a bounded rank polynomial of degree 2, but $2P(x) = |x|/2 \bmod 1$ is an infinite rank polynomial of degree 1. Because of the failure of Theorem 4.1 at the endpoint $k = p$, we will need to require certain polynomials to have degree at least two in our arguments; but we will be able to eliminate all linear polynomials from our analysis by exploiting the freedom to pass to finite index subspaces.

6. MULTILINEAR CONCATENATION

We now begin the proof of Theorem 4.6. The strategy will be to obtain enough control on the biased form $T \in \text{CSM}^{s+1}(V)$ that one can explicitly write this form as $d^{s+1}P$ on a bounded index subspace V' for some bounded rank polynomial $P \in \text{Poly}_{\leq s+1, \text{BR}}(V')$.

To get some intuition as to what expressions such as $d^{s+1}P$ look like, consider the case when $s + 1 = 4$, $V' = V$, and P takes the form $P = QR$ for some quadratic polynomials $Q, R \in \text{Poly}_{\leq 2, \text{BR}}(V \rightarrow \mathbb{F})$. Clearly, P is of bounded rank. A brief computation using the discrete Leibniz rule (D.2) then reveals that

$$\begin{aligned} \Delta_a \Delta_b \Delta_c \Delta_d (QR) &= (\Delta_a \Delta_b Q)(\Delta_c \Delta_d R) + (\Delta_a \Delta_c Q)(\Delta_b \Delta_d R) + (\Delta_a \Delta_d Q)(\Delta_b \Delta_c R) \\ &\quad + (\Delta_b \Delta_c Q)(\Delta_a \Delta_d R) + (\Delta_b \Delta_d Q)(\Delta_a \Delta_c R) + (\Delta_c \Delta_d Q)(\Delta_a \Delta_b R) \end{aligned}$$

for any $a, b, c, d \in V$. Thus, if we let $B, C \in \text{CSM}^2(V \rightarrow \mathbb{F})$ be the quadratic forms $B := d^2Q$, $C := d^2R$, then we have

$$\begin{aligned} d^4P(a, b, c, d) &= B(a, b)C(c, d) + B(a, c)C(b, d) + B(a, d)C(b, c) \\ &\quad + B(b, c)C(a, d) + B(b, d)C(a, c) + B(c, d)C(a, b). \end{aligned}$$

By Lemma 4.5, we thus see that any quadrilinear form $T \in \text{CSM}^4(V)$ of the form

$$\begin{aligned} T(a, b, c, d) &= B(a, b)C(c, d) + B(a, c)C(b, d) + B(a, d)C(b, c) \\ &\quad + B(b, c)C(a, d) + B(b, d)C(a, c) + B(c, d)C(a, b) \end{aligned}$$

for some $B, C \in \text{CSM}^2(V)$ will be of the desired form d^4P for Theorem 4.6.

To generalise this discussion to higher dimensions, we introduce the following operation.

Definition 6.1 (Concatenation). Let $S \in \text{CSM}^k(V)$ and $T \in \text{CSM}^l(V)$ for some standard integers $k, l \geq 1$. We define the *concatenation* $S * T \in \text{CSM}^{k+l}(V)$ of S and T by the formula

$$(S * T)(h_1, \dots, h_{k+l}) = \sum_{\{1, \dots, k+l\} = A \sqcup B} S((h_i)_{i \in A}) T((h_j)_{j \in B})$$

where the sum ranges over all partitions of $\{1, \dots, k+l\}$ into a k -element set A and an l -element set B , and we define $S((h_i)_{i \in A})$ by enumerating A arbitrarily (the precise ordering is not relevant due to the symmetry of S), and similarly for $T((h_j)_{j \in B})$. Thus, for instance, if $S, T \in \text{CSM}^2(V)$, then

$$(S * T)(a, b, c, d) := S(a, b)T(c, d) + S(a, c)T(b, d) + S(a, d)T(b, c) \\ + S(b, c)T(a, d) + S(b, d)T(a, c) + S(c, d)T(a, b).$$

It is not hard to see that $S * T$ is indeed multilinear and symmetric. The fact that it is classical is also easily seen after observing that all the binomial coefficients $\binom{p}{j}$ with $1 \leq j < p$ are divisible by p and thus vanish on \mathbb{F} . The operation $*$ is also easily seen to be bilinear, commutative, and associative. As the previous discussion already indicated, this operation is closely related to multiplication on (classical) polynomials. More precisely, we have:

Lemma 6.2 (Product rule). *Let $k, l \geq 1$ be standard integers. If $P \in \text{Poly}_{\leq k}(V \rightarrow \mathbb{F})$ and $Q \in \text{Poly}_{\leq l}(V \rightarrow \mathbb{F})$, then $d^{k+l}(PQ) = (d^k P) * (d^l Q)$.*

Proof. We apply the discrete Leibniz rule (D.2) repeatedly to expand out the derivative

$$\Delta_{h_1} \dots \Delta_{h_{k+l}}(PQ) \tag{6.1}$$

with $h_1, \dots, h_{k+l} \in V$. Note that if P accepts more than k derivatives, or Q accepts more than l derivatives, then the resulting term in the expansion of (6.1) vanishes. Thus the only terms in (6.1) that survive are those in which P is differentiated exactly k times, and Q differentiated exactly l times. Collecting all such terms one obtains the claim. \square

From this lemma and Lemma 4.5, we see that any form $T \in \text{CSM}^{s+1}(V)$ which can be expressed (possibly after passing to a bounded index subspace) as a linear combination of concatenations $T_1 * \dots * T_k$ of classical symmetric multilinear forms of degree strictly less than $s+1$, will satisfy the conclusions of Theorem 4.6.

In the high characteristic case $p > s+1$, it turns out that these concatenations are the only expressions one needs to consider to establish Theorem 4.6. Unfortunately the situation is more complicated in the low characteristic case $p \leq s+1$. This can be illustrated by using the symmetric polynomial $S_4 \in \text{Poly}_{\leq 4}(\mathbb{F}_2^n \rightarrow \mathbb{F}_2)$ from Example 1.16 (this is on a finite-dimensional space rather than a limit finite-dimensional space, but let us ignore this technicality for this discussion). A routine calculation reveals that the quartilinear form $d^4 S_4 \in \text{CSM}^4(\mathbb{F}_2^n)$ can be expressed as

$$d^4 S_4(a, b, c, d) = B(a, b)B(c, d) + B(a, c)B(b, d) + B(a, d)B(b, c) \tag{6.2}$$

for all $a, b, c, d \in \mathbb{F}_2^n$, where $B \in \text{CSM}^2(\mathbb{F}_2^n)$ is the bilinear form $B = d^2 S_2$, thus

$$B(a, b) := \sum_{1 \leq i, j \leq n: i \neq j} a_i b_j.$$

(The identity (6.2), which was already observed in [33], [22], can be established by testing it on generators $a, b, c, d \in \{e_1, \dots, e_n\}$.) The right-hand side of (6.2) is formally of

the form $(B * B)/2!$, but the operation of dividing by $2!$ is not well-defined in characteristic two, and so in fact one cannot easily express (6.2) in terms of the concatenation operation. Instead, we have to introduce a new operation to handle expressions of this form:

Definition 6.3 (Symmetric power). Let $T \in \text{CSM}^k(V)$ for some standard integers $k \geq 2$ and $m \geq 1$. We define the *symmetric power* $\text{Sym}^m(T) \in \text{CSM}^{mk}(V)$ by the formula

$$\text{Sym}^m(T)(h_1, \dots, h_{mk}) = \sum_{\mathcal{A}} \prod_{A \in \mathcal{A}} T((h_i)_{i \in A}),$$

where the sum ranges over all partitions \mathcal{A} of $\{1, \dots, mk\}$ into m subsets A of cardinality k each. For instance, if $m = k = 2$, then

$$\text{Sym}^2(T)(a, b, c, d) = T(a, b)T(c, d) + T(a, c)T(b, d) + T(a, d)T(b, c).$$

Again, it is clear that $\text{Sym}^k(T)$ is symmetric and multilinear; the fact that it is classical follows by observing that when p of the arguments of $\text{Sym}^k(T)$ are set to be equal, then the multiplicity of each term is a multiple of p (as it is equal to $p!$ divided by a number of factorials that are strictly less than $p!$; note here we use the hypothesis $k \geq 2$). Because of our need to avoid the $k = 1$ case, we will have to take some care to eliminate all linear forms from the arguments in the next section, by using the trick of passing to a finite index subspace to make these forms vanish.

Observe that

$$m! \text{Sym}^m(T) = T * \dots * T$$

where the right-hand side contains m copies of T . Thus, in high characteristic $p > m$, one can write the symmetric power in terms of the concatenation operation by the formula

$$\text{Sym}^m(T) = \frac{T * \dots * T}{m!}. \quad (6.3)$$

However, in the low characteristic case the symmetric power operation cannot be reduced easily to the concatenation operation, and we need to consider the two operations separately.

There is an analogue of Lemma 6.2:

Lemma 6.4 (Symmetric power rule). *Let $k \geq 2$ and $m \geq 1$ be standard integers, and let $T \in \text{CSM}^k(V)$ be a classical symmetric multilinear form. Then there exists $Q \in \text{Poly}_{\leq mk}(V \rightarrow \mathbb{F})$ such $d^{mk}Q = \text{Sym}^m(T)$. Furthermore, if $m \geq 2$, then Q has bounded rank.*

Proof. By Lemma 4.5 we may write $T = d^k P$ for some $P \in \text{Poly}_{\leq k}(V \rightarrow \mathbb{F})$.

By a limiting argument it suffices to establish the claim when V is finite dimensional, as long as the bound in “bounded rank” depends only on p, k, m and not on the dimension of V .

Heuristically, in view of (6.3) and Lemma 6.2, it is natural to try to set Q equal to $P^m/m!$. This works in the high characteristic case $m < p$, but not in the low characteristic case $p \geq m$ due to the non-invertibility of $m!$. To get around this, we will use the binomial coefficient $\binom{P}{m}$ instead of $P^m/m!$; but this requires lifting P to a larger group than $\mathbb{F} = \mathbb{Z}/p\mathbb{Z}$.

We turn to the details. We let $M \geq 0$ be the first integer such that $m < p^{M+1}$, thus $p^M \leq m$. Using Lemma 1.7(v) repeatedly, we may find a polynomial $\tilde{P}_M \in \text{Poly}_{\leq k+M(p-1)}(V \rightarrow \mathbb{T})$ such that $p^M \tilde{P}_M = \iota(P)$. In particular, \tilde{P}_M takes values in the $(p^{M+1})^{\text{th}}$ roots of unity. We may thus pull \tilde{P}_M back to the cyclic group $\mathbb{Z}/p^{M+1}\mathbb{Z}$ to obtain a polynomial $P_M \in \text{Poly}_{\leq k+M(p-1)}(V \rightarrow \mathbb{Z}/p^{M+1}\mathbb{Z})$ such that $P_M = P \bmod p$.

An inspection of the formula $\binom{n}{m} = \frac{n(n-1)\dots(n-m+1)}{m(m-1)\dots 1}$ for a binomial coefficient reveals that the map $n \mapsto \binom{n}{m} \bmod p$ is periodic with period p^{M+1} whenever $m < p^{M+1}$. In particular, by abuse of notation we may define the binomial coefficient $\binom{n}{m} \bmod p \in \mathbb{F}$ whenever $m < p^{M+1}$ and $n \in \mathbb{Z}/p^{M+1}\mathbb{Z}$. We then set $Q := \binom{P_M}{m} \bmod p$.

We first verify that Q is a polynomial of degree $\leq mk$. For inductive reasons, we will prove the more general claim that for any $j \geq 0$, any $h_1, \dots, h_j \in V$, and any $0 \leq m' < p^{M+1}$, the expression $\binom{\Delta_{h_1} \dots \Delta_{h_j} P_M}{m'}$ has degree at most $k - j + (m' - 1) \max(k - j, 1)$. Clearly, this implies the previous claim by setting $j := 0$ and $m' := m$.

We first address the degenerate case when $k - j + (m' - 1) \max(k - j, 1)$ is negative, so in particular $m' \leq j - k$. The polynomial $\Delta_{h_1} \dots \Delta_{h_j} P_M$ has degree $\leq k + M(p-1) - j$, so by Lemma 1.7(v), it is divisible by p^{a+1} whenever $0 \leq a \leq M$ and $k + a(p-1) - j < 0$. In particular, $\Delta_{h_1} \dots \Delta_{h_j} P_M$ is divisible by $p^{\lfloor \frac{m'-1}{p-1} \rfloor + 1}$. On the other hand, observe that $\binom{n}{m'}$ is divisible by p whenever n is divisible by p^a and $m' < p^a$. Since $m' < p^{\lfloor \frac{m'-1}{p-1} \rfloor + 1}$, we obtain the claim.

To handle the non-degenerate cases when $k - j + (m' - 1) \max(k - j, 1) \geq 0$, we use downward induction on j . The claim is vacuously true for j sufficiently large, so we assume inductively that the claim is proven for all larger values of j ; for fixed j , we also assume inductively that the claim is proven for all smaller values of m' . By Lemma 1.7(i), it suffices to show that the expression

$$\Delta_{h_{j+1}} \left(\binom{\Delta_{h_1} \dots \Delta_{h_j} P_M}{m'} \right) \tag{6.4}$$

has degree at most $k - j + (m' - 1) \max(k - j, 1) - 1$ for all $h_{j+1} \in V$.

From the combinatorial identity

$$\binom{n + n'}{m'} = \sum_{i=0}^{m'} \binom{n}{i} \binom{n'}{m' - i}$$

we see that

$$\Delta_h \binom{F}{m'} = \sum_{i=1}^{m'} \binom{\Delta_h F}{i} \binom{F}{m' - i} \quad (6.5)$$

whenever $h \in V$ and $F : V \rightarrow \mathbb{Z}/p^{M+1}\mathbb{Z}$. We may therefore expand (6.4) as

$$\sum_{i=1}^{m'} \binom{\Delta_{h_1} \cdots \Delta_{h_{j+1}} P_M}{i} \binom{\Delta_{h_1} \cdots \Delta_{h_j} P_M}{m' - i}.$$

In each summand, we apply the two induction hypotheses to conclude that the first factor in the summand has degree $\leq k - j - 1 + (i - 1) \max(k - j - 1, 1)$, and the second factor has degree $\leq k - j + (m' - i - 1) \max(k - j, 1)$. A routine computation shows that

$$(k - j - 1 + (i - 1) \max(k - j - 1, 1)) + (k - j + (m' - i - 1) \max(k - j, 1)) \leq k - j + (m' - 1) \max(k - j, 1) - 1$$

whenever $i \geq 1$ (treating the cases $k - j > 1$, $k - j \leq 1$ separately), and the claim then follows from Lemma 1.7(iv). Thus Q has degree $\leq mk$ as desired.

Now we compute the derivative $d^{mk}Q$ of $Q = \binom{P_M}{m}$. Using (6.5) we have

$$\Delta_h \binom{P_M}{m} = \sum_{i=1}^m \binom{\Delta_h P_M}{i} \binom{P_M}{m - i}$$

for any $h \in V$. By the above computations, the polynomial $\binom{\Delta_h P_M}{i} \binom{P_M}{m - i}$ has degree

$$\leq ((k - 1) + (i - 1)(k - 1)) + (k + (m - i - 1)k) = mk - i.$$

In particular, all the terms with $i > 1$ have degree strictly less than $mk - 1$ and thus will not contribute to $d^{mk} \binom{P_M}{m}$. The $i = 1$ term can be simplified as $\Delta_h P \binom{P_M}{m - 1}$. We conclude that

$$d^{mk} \binom{P_M}{m} (h_1, \dots, h_{mk}) = d^{mk-1} ((\Delta_{h_{mk}} P) \binom{P_M}{m - 1}) (h_1, \dots, h_{mk-1}).$$

Expanding this out using Lemma 6.2 we have

$$\begin{aligned} d^{mk} \binom{P_M}{m} (h_1, \dots, h_{mk}) &= (d^{k-1} (\Delta_{h_{mk}} P) * d^{mk-k} \binom{P_M}{m}) (h_1, \dots, h_{mk-1}) \\ &= \sum_{1 \leq i_1 < \dots < i_{k-1} < mk} d^k P(h_{i_1}, \dots, h_{i_{k-1}}, h_{mk}) d^{(m-1)k} \binom{P_M}{m-1} (h_{j_1}, \dots, h_{j_{(m-1)k}}) \end{aligned}$$

where $1 \leq j_1 < \dots < j_{(m-1)k} < mk$ are the ordered enumeration of the set $\{1, \dots, mk - 1\} \setminus \{i_1, \dots, i_{k-1}\}$. The claim $d^{mk} \binom{P_M}{m} = \text{Sym}^m(d^k P)$ then follows by induction on m . \square

Example 6.5. We illustrate the above lemma with $p = m = k = 2$, with $V = \mathbb{F}^n$ and $P = S_2$ the symmetric polynomial from Example 1.16, thus $S_2 : V \rightarrow \mathbb{F}$ is the classical quadratic polynomial

$$S_2(x_1, \dots, x_n) = \sum_{1 \leq i < j \leq n} x_i x_j.$$

The bilinear form $B := d^2 S_2 : V^2 \rightarrow \mathbb{F}$ is then given as

$$B((x_1, \dots, x_n), (y_1, \dots, y_n)) = \sum_{1 \leq i, j \leq n: i \neq j} x_i y_j;$$

this is a classical symmetric bilinear form.

Lemma 6.4 asserts the existence of a classical quartic $Q : V \rightarrow \mathbb{F}$ with

$$d^4 Q(a, b, c, d) = B(a, b)B(c, d) + B(a, c)B(b, d) + B(a, d)B(b, c). \quad (6.6)$$

By (6.2) we see that we can take $Q = S_4$. To connect this with the proof of Lemma 6.4, we recall that $P = S_2 = L \bmod 2$, so if we set $P_1 := L \bmod 4$, then $P = P_1 \bmod 2$, and $S_4 = \binom{P_1}{2} \bmod 2$.

Theorem 4.6 now follows from a more explicit claim:

Theorem 6.6 (Explicit inverse Gowers for classical symmetric multilinear forms). *Let $s \geq 2$ be such that $\text{GIP}(s')$ is true for all $0 \leq s' < s$, let V be limit finite-dimensional, and let $T \in \text{CSM}^{s+1}(V)$ be such that $|\mathbb{E}_{h_1, \dots, h_{s+1} \in V} e(\iota(T(h_1, \dots, h_{s+1})))| \gg 1$. Then there exists a bounded index subspace V' of V such that on $(V')^{s+1}$, T is a linear combination (over \mathbb{F}) of a bounded number of expressions of the form*

$$\text{Sym}^{m_1}(S_1) * \dots * \text{Sym}^{m_r}(S_r) \quad (6.7)$$

for some $m_1, \dots, m_r \geq 1$ and $2 \leq k_1, \dots, k_r < s+1$ and $S_i \in \text{CSM}^{k_i}(V')$ for $i = 1, \dots, r$ with

$$m_1 k_1 + \dots + m_r k_r = s + 1.$$

Indeed, by repeatedly applying Lemma 4.5 we find that $S_i = d^{k_i} P_i$ for some $P_i \in \text{Poly}_{\leq k_i}(V \rightarrow \mathbb{F})$. By Lemma 6.4 $\text{Sym}^{m_i}(d^{k_i} P_i) = d^{k_i m_i} Q_i$ for some $Q_i \in \text{Poly}_{\leq m_i k_i}(V \rightarrow \mathbb{F})$, and Q_i is of bounded rank if $m_i \geq 2$. Now by Lemma 6.2, we see that any expression of the form (6.7) can be expressed as $d^{s+1} Q$ on V' where $Q = \prod_i Q_i \in \text{Poly}_{\leq s+1, \text{BR}}(V' \rightarrow \mathbb{F})$ (note that the product of at least two polynomials of degree ≥ 1 is necessarily of bounded rank), and Theorem 4.6 follows by linearity.

It remains to establish Theorem 6.6. To illustrate the type of result one is seeking here, in the case $s+1 = 6$, one has a classical symmetric sextilinear form $T \in \text{CSM}^6(V)$ which is biased in the sense that

$$|\mathbb{E}_{a, b, c, d, e, f \in V} e(\iota(T(a, b, c, d, e, f)))| \gg 1,$$

and one wishes to conclude that on a bounded index subspace V' of V , $T(a, b, c, d, e, f)$ can be decomposed into a bounded number of pieces such as the expression

$$B(a, b)B(c, d)B(e, f)$$

plus $\frac{6!}{2!2!2!3!} - 1 = 14$ other permutations (for some $B \in \text{CSM}^2(V')$), adding up to $\text{Sym}^3(B)$, or

$$B(a, b)B(c, d)B'(e, f)$$

plus $\frac{6!}{2!2!2!2!} - 1 = 44$ other permutations (for some $B, B' \in \text{CSM}^2(V')$), adding up to $\text{Sym}^2(B) * B'$, or

$$C(a, b, c)C(d, e, f)$$

plus $\frac{6!}{3!3!2!} - 1 = 9$ other permutations (for some $C \in \text{CSM}^3(V')$), adding up to $\text{Sym}^2(C)$, or

$$B(a, b)D(c, d, e, f)$$

plus $\frac{6!}{2!4!} - 1 = 14$ other permutations (for some $B \in \text{CSM}^2(V')$ and $D \in \text{CSM}^4(V')$), adding up to $B * D$.

7. EQUIDISTRIBUTION OF MULTILINEAR MAPS

In order to establish Theorem 6.6 (and thus Theorem 4.6) we will need an equidistribution theory for classical symmetric multilinear maps, analogous to that in [22], [28]. We introduce some definitions:

Definition 7.1 (Bounded rank for multilinear forms). Let $k \geq 1$ and let V be limit finite-dimensional. A form $T \in \text{CSM}^k(V)$ is said to be *bounded rank* if there exist a bounded number of forms $S_i \in \text{CSM}^{k_i}(V)$, $i = 1, \dots, m$ for some $1 \leq k_i < k$ such that for $h_1, \dots, h_k \in V$, the expression $T(h_1, \dots, h_k)$ is a function of expressions of the form $S_i(h_{j_1}, \dots, h_{j_{k_i}})$ for some $1 \leq i \leq m$ and $1 \leq j_1 < \dots < j_{k_i} \leq k$, and *unbounded rank* otherwise. The space of bounded rank forms will be denoted $\text{CSM}_{\text{BR}}^k(V)$; it is clearly a subspace of $\text{CSM}^k(V)$.

Thus, for instance, if $k = 4$ and T takes the form

$$T(a, b, c, d) = B(a, b)B(c, d) + B(a, c)B(b, d) + B(a, d)B(b, c)$$

(i.e. $T = \text{Sym}^2(B)$) for some $B \in \text{CSM}^2(V)$, then T would be bounded rank.

Our starting point will be the following result of Kaufman and Lovett [28] (which in turn is based on the earlier paper [22]), which links bounded rank with bias:

Proposition 7.2 (Bias criterion). *Let $k \geq 1$ be a standard integer, and let V be limit finite-dimensional. Then a form $T \in \text{CSM}^k(V)$ is bounded rank if and only if*

$$|\mathbb{E}_{h_1, \dots, h_k \in V} e(\iota(T(h_1, \dots, h_k)))| \gg 1.$$

Proof. The “only if” part is easy: if T is bounded rank, then by Fourier analysis, the function $e(\iota(T(h_1, \dots, h_k)))$ is a bounded linear combination of functions of the form $e(\iota(\sum_{i=1}^m c_i S_i(h_{j_1}, \dots, h_{j_{k_i}})))$ for $c_i \in \mathbb{F}$. Such functions can be factorised as $\prod_{j=1}^d f_j(h_1, \dots, h_k)$ where the f_j are bounded functions not depending on h_j . On the other hand, we clearly have

$$\mathbb{E}_{h_1, \dots, h_k \in V} e(\iota(T(h_1, \dots, h_k))) \overline{e(\iota(T(h_1, \dots, h_k)))} = 1.$$

Applying the pigeonhole principle we conclude that

$$|\mathbb{E}_{h_1, \dots, h_k \in V} e(\iota(T(h_1, \dots, h_k))) \prod_{j=1}^d \overline{f_j(h_1, \dots, h_k)}| \gg 1$$

for at least one collection f_1, \dots, f_d of bounded functions; the claim then follows from Lemma 1.17(iv) (and Lemma 4.5).

We now turn to the “if” part. This result follows easily from [28, Theorem 2]. Indeed, applying that theorem, we see that $T(h_1, \dots, h_k)$ is a function of a bounded number of polynomials $\Delta_v T(h_1, \dots, h_k)$ for $v \in V^k$, but by the multilinearity of T we see that all such derivatives are functions of expressions of the form $S_i(h_{j_1}, \dots, h_{j_{k_i}})$ for some $1 \leq j_1 < \dots < j_{k_i} \leq k$, as required. \square

Using the above proposition, we can obtain a criterion for joint equidistribution for certain systems of CSM forms; the precise definition of joint equidistribution is given in Definition C.20. We need another definition:

Definition 7.3 (Systems of CSM forms). Let $k_0 \geq 1$ be standard, and let V be limit finite-dimensional. A CSM-system $T = (T_{k,i})_{1 \leq k < k_0; 1 \leq i \leq m_k}$ of degree $< k_0$ is a collection of forms $T_{k,i} \in \text{CSM}^k(V)$ for $1 \leq k < k_0$ and $1 \leq i \leq m_k$, where m_1, \dots, m_{k_0-1} are standard natural numbers. A CSM-system is said to be *regular* if, for each $1 \leq k < k_0$, the forms $T_{k,1}, \dots, T_{k,m_k} \in \text{CSM}^k(V)$ are linearly independent modulo $\text{CSM}_{\text{BR}}^k(V)$, thus one has

$$a_1 T_{k,1} + \dots + a_k T_{k,m_k} \notin \text{CSM}_{\text{BR}}^k(V)$$

whenever $a_1, \dots, a_k \in \mathbb{F}$ are not all zero.

Let $S \in \text{CSM}^d(V)$ be a classical symmetric multilinear form of some (standard) degree $d \geq 1$. We say that S is *measurable* with respect to a CSM-system $T = (T_{k,i})_{1 \leq k < k_0; 1 \leq i \leq m_k}$ if there is a functional relationship of the form

$$S = F((T_\alpha)_{\alpha \in A})$$

where A is the set of tuples

$$\alpha = (k_\alpha, i_\alpha, j_{\alpha,1}, \dots, j_{\alpha,k_\alpha})$$

with $1 \leq k_\alpha < k_0$, $1 \leq i_\alpha \leq m_{k_\alpha}$, and $1 \leq j_{\alpha,1} < j_{\alpha,2} < \dots < j_{\alpha,k_\alpha} \leq d$ (in particular, this forces $k_\alpha \leq d$), and $F : \mathbb{F}^A \rightarrow \mathbb{F}$ is a function.

Example 7.4. A collection $L_1, \dots, L_{m_1} : V \rightarrow \mathbb{F}$ of linear forms, together with a collection $B_1, \dots, B_{m_2} : V^2 \rightarrow \mathbb{F}$ of classical symmetric bilinear forms, will form a CSM-system of degree < 3 . In order for this system to be regular, the linear forms L_1, \dots, L_{m_1} must be linearly independent, and no non-trivial linear combination of the B_1, \dots, B_{m_2} can be of bounded rank (i.e. expressible in terms of boundedly many linear forms). A quartilinear form such as $B_1 * B_2$, $B_1 * L_1 * L_2$, $B_1 * \text{Sym}^2(L_1)$, $\text{Sym}^2(B_1)$, or linear combinations thereof, will be measurable with respect to T (provided that m_1, m_2 are large enough so that these expressions make sense, of course).

Proposition 7.2 can now be recast as follows:

Corollary 7.5 (Bias criterion, again). *Let $k \geq 1$ be a standard integer, and let V be limit finite-dimensional. Let $T \in \text{CSM}^k(V)$. Then one has*

$$|\mathbb{E}_{h_1, \dots, h_k \in V} e(\iota(T(h_1, \dots, h_k)))| \gg 1$$

if and only if T is measurable with respect to a CSM-factor $S = (S_{k',i})_{1 \leq k' < k; 1 \leq i \leq m_{k'}}$ of degree $< k$.

The CSM-factor S given by the above corollary is not necessarily regular, but we may always regularise it as follows:

Lemma 7.6 (Regularity lemma). *Let $k_0 \geq 1$ be a standard integer, and let $T = (T_{k,i})_{1 \leq k < k_0; 1 \leq i \leq m_k}$ be a CSM-factor of degree $< k_0$. Then there exists a regular CSM-factor $S = (S_{k,i})_{1 \leq k < k_0; 1 \leq i \leq m'_k}$ of degree $< k_0$, such that every multilinear form $T_{k,i}$ in T is measurable with respect to S .*

Proof. We assume inductively that the claim has already been proven for all smaller values of k_0 (this hypothesis is vacuous for $k_0 = 1$).

We consider the top-order forms $T_{k_0-1,1}, \dots, T_{k_0-1,m_{k_0-1}} \in \text{CSM}^{k_0-1}(V)$, projected to the quotient space $\text{CSM}^{k_0-1}(V) / \text{CSM}_{\text{BR}}^{k_0-1}(V)$. As every finitely generated vector space has a finite basis, we may thus find forms $S_{k_0-1,1}, \dots, S_{k_0-1,m'_{k_0-1}} \in \text{CSM}^{k_0-1}(V)$ that are linearly independent modulo $\text{CSM}_{\text{BR}}^{k_0-1}(V)$, thus that each $T_{k_0-1,i}(h_1, \dots, h_k)$ is a linear combination (over \mathbb{F}) of the $S_{k_0-1,1}(h_1, \dots, h_k), \dots, S_{k_0-1,m'_{k_0-1}}(h_1, \dots, h_k)$, plus a bounded rank form of h_1, \dots, h_k , which by definition of bounded rank can be expressed as a function of a bounded family of forms $U(h_{i_1}, \dots, h_{i_{k'}})$ with $1 \leq k' < k$, $U \in \text{CSM}^{k'}(V)$, and $1 \leq i_1 < \dots < i_{k'} \leq k'$. We may add all such forms U to the list of lower order forms $T_{k',1}, \dots, T_{k',m_{k'}}$. Applying the induction hypothesis to those lower order forms we then obtain the claim. \square

Applying this regularity lemma to Corollary 7.5 we obtain the following improvement:

Corollary 7.7 (Regularised bias criterion). *Let $k \geq 1$ be a standard integer, and let V be limit finite-dimensional. Let $T \in \text{CSM}^k(V)$. Then one has*

$$|\mathbb{E}_{h_1, \dots, h_k \in V} e(\iota(T(h_1, \dots, h_k)))| \gg 1$$

if and only if T is measurable with respect to a regular CSM-factor $S = (S_{k',i})_{1 \leq k' < k; 1 \leq i \leq m_{k'}}$ of degree $< k$.

To use this criterion, we need the following *counting lemma* which complements the regularity lemma.

Lemma 7.8 (Counting lemma). *Let $k_0 \geq 1$ be a standard integer, and let $T = (T_{k,i})_{1 \leq k < k_0; 1 \leq i \leq m_k}$ be a regular CSM-system. Then for any $d \geq 1$, the expressions*

$$T_{k,i,j_1, \dots, j_k} : (h_1, \dots, h_d) \mapsto T_{k,i}(h_{j_1}, \dots, h_{j_k}),$$

where $1 \leq k < k_0$, $1 \leq i \leq m_k$, and $1 \leq j_1 < \dots < j_k \leq d$, as functions from V^d to \mathbb{F} , are jointly equidistributed (as defined in Definition C.20).

Proof. By the Weyl equidistribution criterion (Lemma C.23), it suffices to show that

$$\mathbb{E}_{h_1, \dots, h_d \in V} e(\iota(\sum_{1 \leq k < k_0} \sum_{i=1}^{m_k} \sum_{1 \leq j_1 < \dots < j_k \leq d} c_{k,i,j_1, \dots, j_k} T_{k,i}(h_{j_1}, \dots, h_{j_k}))) = o(1) \quad (7.1)$$

whenever $c_{k,i,j_1, \dots, j_k} \in \mathbb{F}$ are not all zero.

Let $1 \leq k_1 < k_0$ be the largest k for which there is a non-zero coefficient c_{k,i,j_1,\dots,j_k} . By relabeling we may assume that it is $c_{k_1,1,1,\dots,k_1}$ which is non-zero. We may then factorise the left-hand side of (7.1) as

$$\mathbb{E}_{h_1,\dots,h_d \in V} e(\iota(T_*(h_1, \dots, h_{k_1}))) \prod_{i=1}^{k_1} f_i(h_1, \dots, h_d)$$

where $f_i : V \rightarrow {}^*\mathbb{C}$ are limit functions bounded in magnitude by 1 which are independent of the h_i variable, and $T_* : V^{k_1} \rightarrow \mathbb{F}$ is the multilinear form

$$T_*(h_1, \dots, h_{k_1}) := \sum_{i=1}^{m_1} c_{k_1,i,1,\dots,k_1} T_{k_1,i}(h_1, \dots, h_{k_1}).$$

By hypothesis, T_* has unbounded rank. By Proposition 7.2, we conclude that

$$\mathbb{E}_{h_1,\dots,h_d \in V} e(\iota(T_*(h_1, \dots, h_{k_1}))) = o(1).$$

Using Lemma 1.17 (and Lemma 4.5) we conclude that

$$\mathbb{E}_{h_1,\dots,h_d \in V} e(\iota(T_*(h_1, \dots, h_{k_1}))) \prod_{i=1}^{k_1} f_i(h_1, \dots, h_d) = o(1)$$

and the claim follows. \square

Remark 7.9. The counting lemma is essentially asserting that a regular CSM-factor is equidistributed on cubes in the sense of Definition C.21, but to formalise this rigorously, one needs to develop a theory of polynomial maps in several variables, that generalises the material in Section C to groups filtered by \mathbb{N}^d rather than \mathbb{N} . This can be done (see [24]), but we will not introduce this additional notation here.

In the next section we will use the counting and regularity lemmas to finish off the proof of Theorem 6.6 and thus Theorem 4.6.

8. CONCLUSION OF THE MULTILINEAR INVERSE CONJECTURE

We are now ready to complete the proof of Theorem 6.6. Let s, T, V be as in that theorem. Applying Corollary 7.7, we conclude that we can write

$$T = F((S_\alpha)_{\alpha \in A}),$$

where $(S_{k,i})_{1 \leq k \leq s; 1 \leq i \leq m_k}$ is a regular CSM-system, A is the set of tuples

$$\alpha = (k_\alpha, i_\alpha, j_{\alpha,1}, \dots, j_{\alpha,k_\alpha})$$

with $1 \leq k_\alpha \leq s$, $1 \leq i_\alpha \leq m_{k_\alpha}$, and $1 \leq j_{\alpha,1} < j_{\alpha,2} < \dots < j_{\alpha,k_\alpha} \leq s+1$, $F : \mathbb{F}^A \rightarrow \mathbb{F}$ is a function, and for each $\alpha \in A$, $S_\alpha \in \text{CSM}^{k_\alpha}(V)$ is the form

$$S_\alpha(h_1, \dots, h_{k_\alpha}) := S_{k_\alpha, i_\alpha}(h_{j_{\alpha,1}}, \dots, h_{j_{\alpha,k_\alpha}}).$$

The linear forms $S_{1,1}, \dots, S_{1,m_1}$ in the regular CSM-system can be eliminated by observing that they are simultaneously constant on some bounded index subspace of V , and so by passing to that subspace (using Lemma 4.7) we may assume that $m_1 = 0$,

i.e. all forms $S_{k,i}$ in the CSM-system are bilinear or higher in order. Of course, the CSM-system remains regular after doing so.

For each S_α , we refer to the set $J_\alpha := \{j_{\alpha,1}, \dots, j_{\alpha,k_\alpha}\} \subset \{1, \dots, s\}$ as the *support* of α .

The permutation group $\text{Sym}(\{1, \dots, s+1\})$ acts on the variables h_1, \dots, h_{s+1} , and thus permutes the index set A . Because T is symmetric (and $(S_\alpha)_{\alpha \in A} : V \rightarrow \mathbb{F}^A$ is surjective, by Lemma 7.8), we see that F is symmetric with respect to this action.

Next, we show that F is also multilinear:

Proposition 8.1. *$F : \mathbb{F}^A \rightarrow \mathbb{F}$ is a linear combination (over \mathbb{F}) of monomials*

$$(x_\alpha)_{\alpha \in A} \mapsto x_{\alpha_1} \dots x_{\alpha_r} \quad (8.1)$$

where $\alpha_1, \dots, \alpha_r$ are elements of A whose supports $J_{\alpha_1}, \dots, J_{\alpha_r}$ partition $\{1, \dots, s+1\}$.

For instance, if $s+1 = 4$, this proposition asserts that $T(a, b, c, d)$ is a linear combination of expressions such as

$$B(a, b)B'(c, d)$$

for bilinear forms $B, B' \in \text{CSM}^2(V)$ in the CSM-system, but not expressions such as

$$B(a, b)B'(a, c)$$

or

$$B(a, b)^2 B(c, d).$$

Proof. Split $A = A_{\ni s+1} \cup A_{\not\ni s+1}$, where $A_{\ni s+1}$ consist of those $\alpha \in A$ whose support J_α contains $s+1$, and $A_{\not\ni s+1} := A \setminus A_{\ni s+1}$. We split $\mathbb{F}^A = \mathbb{F}^{A_{\ni s+1}} \times \mathbb{F}^{A_{\not\ni s+1}}$ in the obvious manner. We claim the linearity statement

$$F(x_{\ni s+1} + y_{\ni s+1}, x_{\not\ni s+1}) = F(x_{\ni s+1}, x_{\not\ni s+1}) + F(y_{\ni s+1}, x_{\not\ni s+1}) \quad (8.2)$$

whenever $x_{\ni s+1}, y_{\ni s+1} \in \mathbb{F}^{A_{\ni s+1}}$ and $x_{\not\ni s+1} \in \mathbb{F}^{A_{\not\ni s+1}}$.

To prove (8.2), it suffices by the linearity of

$$T = F((S_\alpha)_{\alpha \in A_{\ni s+1}}, (S_\alpha)_{\alpha \in A_{\not\ni s+1}})$$

in the h_{s+1} variable to locate $h_1, \dots, h_s, h_{s+1}, h'_{s+1} \in V$ such that

$$S_{\ni s+1}(h_1, \dots, h_s, h_{s+1}) = x_{\ni s+1}$$

$$S_{\ni s+1}(h_1, \dots, h_s, h'_{s+1}) = y_{\ni s+1}$$

$$S_{\not\ni s+1}(h_1, \dots, h_s, h_{s+1}) = x_{\not\ni s+1}$$

since this implies that

$$S_{\ni s+1}(h_1, \dots, h_s, h_{s+1} + h'_{s+1}) = x_{\ni s+1} + y_{\ni s+1}$$

and

$$S_{\not\ni s+1}(h_1, \dots, h_s, h'_{s+1}) = x_{\not\ni s+1}.$$

But the existence of $h_1, \dots, h_s, h_{s+1}, h'_{s+1} \in V$ with these properties follows immediately from Lemma 7.8.

By symmetry, (8.2) generalises to

$$F(x_{\ni j} + y_{\ni j}, x_{\not\ni j}) = F(x_{\ni j}, x_{\not\ni j}) + F(y_{\ni j}, x_{\not\ni j}) \quad (8.3)$$

whenever $1 \leq j \leq s+1$, $x_{\ni j}, y_{\ni j} \in \mathbb{F}^{A_{\ni j}}$ and $x_{\not\ni j} \in \mathbb{F}^{A_{\not\ni j}}$, where $A_{\ni j}, A_{\not\ni j}$ are defined analogously to $A_{\ni s+1}, A_{\not\ni s+1}$ by replacing $s+1$ with j , and \mathbb{F}^A is identified with $\mathbb{F}^{A_{\ni j}} \times \mathbb{F}^{A_{\not\ni j}}$ in the obvious manner. We claim that the identities (8.3) imply that F is a linear combination of the monomials (8.1) (note that the converse claim is clear).

To establish this implication, we induct on s . The case $s = 0$ is easily verified, so suppose $s \geq 1$ and the claim has already been proven for smaller s . For every $\alpha \in A_{\ni s+1}$, we consider the derivative $\Delta_{e_\alpha} F : \mathbb{F}^A \rightarrow \mathbb{F}$ of F in the basis direction e_α . From (8.3) (applied to each j in the support of α) we see that $\Delta_{e_\alpha} F(x)$ is independent of any coefficient x_β whose support J_β intersects J_α , and thus descends to a function F_α on $\mathbb{F}^{A_{J_\alpha}^\perp}$, where $A_{J_\alpha}^\perp$ is the set of $\beta \in A$ whose support lies in $\{1, \dots, s+1\} \setminus J_\alpha$. The linearity properties (8.3) for $j \notin J_\alpha$ descend from F to F_α , so by the induction hypothesis (and relabeling) each F_α is a linear combination of monomials (8.1) with $J_{\alpha_1}, \dots, J_{\alpha_r}$ partitioning $\{1, \dots, k\} \setminus J_\alpha$. In particular, $x_\alpha F_\alpha$ is a linear combination of monomials of the desired form. The function $F - \sum_{\alpha \in A_{\ni s+1}} x_\alpha F_\alpha$ is then invariant in the e_α direction for all $\alpha \in A_{\ni s+1}$ and thus vanishes by (8.3) (which implies in particular that $F(0, x_{\not\ni s+1}) = 0$ for all $x_{\not\ni s+1} \in \mathbb{F}^{A_{\not\ni s+1}}$). The claim follows. \square

Remark 8.2. Proposition 8.1 can be viewed as a variant of Corollary C.22, but to make this connection precise one would have to generalise the machinery in Appendix C to \mathbb{N}^s -filtered groups, as in [24].

From Proposition 8.1 we have

$$F((x_\alpha)_{\alpha \in A}) = \sum_{\{\alpha_1, \dots, \alpha_r\}} c_{\{\alpha_1, \dots, \alpha_r\}} x_{\alpha_1} \dots x_{\alpha_r}$$

for some coefficients $c_{\{\alpha_1, \dots, \alpha_r\}}$, where $\{\alpha_1, \dots, \alpha_r\}$ ranges over all unordered collections of elements $\alpha_1, \dots, \alpha_r$ of A whose supports partition $\{1, \dots, s+1\}$. The multilinear forms $x_{\alpha_1} \dots x_{\alpha_r}$ are clearly linearly independent, and so the coefficients $c_{\{\alpha_1, \dots, \alpha_r\}}$ are uniquely determined by F . In particular, since F is symmetric with respect to the permutation action of $\text{Sym}(\{1, \dots, s\})$, the coefficients $c_{\{\alpha_1, \dots, \alpha_r\}}$ must be symmetric also. Substituting $x_\alpha := S_\alpha$, we now write

$$T = \sum_{\{\alpha_1, \dots, \alpha_r\}} c_{\{\alpha_1, \dots, \alpha_r\}} S_{\alpha_1} \dots S_{\alpha_r}.$$

From the symmetry of the coefficients $c_{\{\alpha_1, \dots, \alpha_r\}}$, we may split this sum into orbits of the action of the permutation group, and conclude that T is in fact a linear combination of the basic symmetric monomials

$$\text{Sym}^{m_1}(S_{k_1, i_1}) * \dots * \text{Sym}^{m_l}(S_{k_l, i_l})$$

with $m_1 k_1 + \dots + m_l k_l = s+1$. Note that all the k_1, \dots, k_l are at least 2, because we have deleted all the linear forms from the CSM-system. This gives Theorem 6.6 as required.

9. A REGULARITY LEMMA AND EQUIDISTRIBUTION FOR NON-CLASSICAL POLYNOMIALS

It remains to establish Theorem 4.2. To do this, it is convenient for inductive reasons to establish a technical strengthening of Theorem 4.2. We first need an analogue of the notion of a regular CSM-system, but now for polynomials instead of multilinear forms:

Definition 9.1 (Regular factor). Let V be a limit finite-dimensional vector space. A *factor* is a bounded family $P = (P_{i,j})_{1 \leq i \leq m; 0 \leq j \leq J_i}$ of polynomials $P_{i,j} \in \text{Poly}_{\leq D_i + j(p-1)}(V \rightarrow {}^*\mathbb{T})$, where $m \geq 0$ and $J_1, \dots, J_m \geq 0$ and $D_1, \dots, D_m \geq 2$ are standard natural numbers, obeying the relations

$$pP_{i,j} = P_{i,j-1} \quad (9.1)$$

for all $1 \leq i \leq m$ and $0 \leq j \leq J_i$ (with the convention $P_{i,-1} = 0$). In particular, each $P_{i,j}$ takes values in the $(p^{j+1})^{\text{th}}$ roots of unity $\frac{1}{p^{j+1}}\mathbb{Z}/\mathbb{Z}$. We refer to m as the *dimension* of the factor, D_1, \dots, D_m as the *initial degrees*, and J_1, \dots, J_m as the *depths*. The *degree* of the factor is the quantity $\sup_{1 \leq i \leq m} D_i + J_i(p-1)$.

A *depth extension* of P is a factor of the form $P' = (P_{i,j})_{1 \leq i \leq m; 0 \leq j \leq J'_i}$, where $J'_i \geq J_i$ for each $1 \leq i \leq m$, and the polynomials $P_{i,j}$ in P' agree with their counterparts in P for $j \leq J_i$. If P' is a depth extension of P , we call P a *depth retraction* of P' . If d is a standard integer, we define the *degree $\leq d$ depth retraction* $P_{\leq d}$ of P to be the retraction formed by deleting all $P_{i,j}$ with $D_i + J_i(p-1) > d$.

If, for every standard integer $k \geq 2$, the polynomials $P_{i,j}$ with $D_i + j(p-1) = k$ are linearly independent in $\text{Poly}_{\leq k}(V \rightarrow {}^*\mathbb{T})$ modulo $\text{Poly}_{\leq k, \text{BR}}(V \rightarrow {}^*\mathbb{T})$, we say that the factor P is *regular*.

A function $Q : V \rightarrow {}^*\mathbb{T}$ is said to be *measurable* with respect to the factor P if one has $Q = F(P_{1,J_1}, \dots, P_{m,J_m})$ for some function $F : {}^*\mathbb{T}^m \rightarrow {}^*\mathbb{T}$.

When working exclusively with classical polynomials, one can set all the depths J_i to zero, and the notion of a factor and a regular factor then become essentially the same as those considered in [22] (see also [28] for a variant of the notion of regularity in the low characteristic case). However, when trying to regularise non-classical polynomials, one unfortunately needs to consider factors of positive depth, which are more technical to study.

Note that in our definition of a regular factor, the degrees D_1, \dots, D_m are at least two. This is because we can eliminate any linear polynomials that arise in the analysis by passing to a finite index subspace. We will need to eliminate the linear case in order to avoid the $k = p$ case of Corollary 5.2, which is false.

Example 9.2. Suppose one has three polynomials $P_{1,0} \in \text{Poly}_{\leq D_1}(V \rightarrow \frac{1}{p}\mathbb{Z}/\mathbb{Z})$, $P_{1,1} \in \text{Poly}_{\leq D_1+p-1}(V \rightarrow \frac{1}{p^2}\mathbb{Z}/\mathbb{Z})$, and $P_{2,0} \in \text{Poly}_{\leq D_2+p-1}(V \rightarrow \frac{1}{p}\mathbb{Z}/\mathbb{Z})$ for some natural numbers $D_1, D_2 \geq 2$, with $pP_{1,1} = P_{1,0}$, then $P = (P_{1,0}, P_{1,1}, P_{2,0})$ would be a factor of dimension 2, initial degrees D_1, D_2 , depths 1, 0, and degree $\max(D_1 + p - 1, D_2)$. One

can view this factor as a map $x \mapsto (P_{1,1}(x), P_{2,0}(x))$ from V from $(\frac{1}{p^2}\mathbb{Z}/\mathbb{Z}) \times (\frac{1}{p}\mathbb{Z}/\mathbb{Z})$ (the polynomial $P_{1,0}$ can be omitted from this map as it is determined by $P_{1,1}$).

From Lemma 1.7(iv), one can find roots $P_{1,2} \in \text{Poly}_{\leq D_1+2(p-1)}(V \rightarrow \frac{1}{p^3}\mathbb{Z}/\mathbb{Z})$ and $P_{2,1} \in \text{Poly}_{\leq D_2+(p-1)}(V \rightarrow \frac{1}{p^2}\mathbb{Z}/\mathbb{Z})$ of $P_{1,1}$ and $P_{2,0}$ respectively (thus $pP_{1,2} = P_{1,1}$ and $pP_{2,1} = P_{2,0}$), then $P' = (P_{1,0}, P_{1,1}, P_{1,2}, P_{2,0}, P_{2,1})$ is a depth extension of the factor $P = (P_{1,0}, P_{1,1}, P_{2,0})$ (or equivalently, P is a depth retraction of P'), in which the depths have been increased from $J_1 = 1, J_2 = 0$ to $J'_1 = 2, J'_2 = 1$. Of course, one can iterate this procedure and perform depth extensions of P to arbitrary depths. If P is interpreted as a map from V to $(\frac{1}{p^2}\mathbb{Z}/\mathbb{Z}) \times (\frac{1}{p}\mathbb{Z}/\mathbb{Z})$, one can view P' as a lift of that map to $(\frac{1}{p^3}\mathbb{Z}/\mathbb{Z}) \times (\frac{1}{p^2}\mathbb{Z}/\mathbb{Z})$, with the original map factoring through the map $(x, y) \mapsto (px, py)$ from $(\frac{1}{p^3}\mathbb{Z}/\mathbb{Z}) \times (\frac{1}{p^2}\mathbb{Z}/\mathbb{Z})$ to $(\frac{1}{p^2}\mathbb{Z}/\mathbb{Z}) \times (\frac{1}{p}\mathbb{Z}/\mathbb{Z})$.

For sake of concreteness, let us now suppose that $D_2 = D_1 + p - 1$. Then P is regular precisely when the degree D_1 polynomial $P_{1,0}$ has unbounded rank, and all non-trivial linear combinations of the degree $D_2 = D_1 + p - 1$ polynomials $P_{1,1}, P_{2,0}$ have unbounded rank. In order for P' to be regular, one must also add the additional property that no non-trivial linear combination of the degree $D_2 + p - 1 = D_1 + 2(p - 1)$ polynomials $P_{1,2}, P_{2,1}$ have unbounded rank. In this case, P is the degree $\leq D_2$ depth retraction of P' .

A convenient property of depth extensions is that they preserve regularity:

Lemma 9.3 (Depth extensions preserve regularity). *Let $P = (P_{i,j})_{1 \leq i \leq m; 0 \leq j \leq J_i}$ be a regular factor, and let $P' = (P'_{i,j})_{1 \leq i \leq m'; 0 \leq j \leq J'_i}$ be a depth extension of P . Then P' is also regular.*

Proof. By induction, it suffices to verify the claim in the case when $J'_i \leq J_i + 1$.

Suppose for contradiction that P' is not regular, then there exists $k \geq 0$ such that the polynomials $P'_{i,j}$ with $D_i + j(p - 1) = k$ have a non-trivial linear dependence in $\text{Poly}_{\leq k}(V \rightarrow {}^*\mathbb{T})$ modulo $\text{Poly}_{\leq k, \text{BR}}(V \rightarrow {}^*\mathbb{T})$, thus

$$\sum_{1 \leq i \leq m; 0 \leq j \leq J_i; D_i + j(p-1) = k} c_{i,j} P'_{i,j} \in \text{Poly}_{\leq k, \text{BR}}(V \rightarrow {}^*\mathbb{T})$$

for some coefficients $c_{i,j} \in \mathbb{Z}$, not all zero.

Suppose that $c_{i,j}$ vanished whenever $j = J_i + 1$, then this linear dependence already occurred in P , implying that P was also not regular, a contradiction. Thus we have $c_{i, J_i+1} \neq 0$ for at least one i ; since $D_i \geq 2$, this also forces $k > p$. (It is here that we crucially need to prevent D_i from being equal to 1.) We now multiply the above linear dependence by p using Corollary 5.2 and (9.1) to conclude that

$$\sum_{1 \leq i \leq m; 0 \leq j \leq J_i; D_i + j(p-1) = k} c_{i,j} P'_{i,j-1} \in \text{Poly}_{\leq k-p+1, \text{BR}}(V \rightarrow {}^*\mathbb{T}).$$

But this is again a non-trivial linear dependence in P , again yielding contradiction. \square

We now localise the exact roots property from Theorem 4.2 to depth extensions of a regular factor. More precisely, we introduce the following property:

Definition 9.4 (Exact roots property). Let s be a standard integer. We say that the *exact roots property* $\text{ER}(s)$ holds if, whenever $P = (P_{i,j})_{1 \leq i \leq m; 0 \leq j \leq J_i}$ is a regular factor on a limit finite-dimensional vector space V of some initial degrees D_1, \dots, D_m , and $Q \in \text{Poly}_{\leq s+1}(V \rightarrow {}^*\mathbb{T})$ is a function of P , and $P' = (P_{i,j})_{1 \leq i \leq m; 0 \leq j \leq J'_i}$ is a depth extension of P with $D_i + J'_i(p-1) > s$ for all $1 \leq i \leq m$, then there exists $R \in \text{Poly}_{\leq s+p}(V' \rightarrow {}^*\mathbb{T})$ that is a function of P' such that $pR = Q$ on V . Furthermore, R can be taken to be a linear combination of those $P_{i,j}$ in P' with $D_i + j(p-1) = s+p$, plus a function of those $P_{i,j}$ with $D_i + j(p-1) < s+p$.

Example 9.5. We continue Example 9.2, again supposing that $D_2 = D_1 + p - 1$ for concreteness. Assume that P is regular. Let $s < D_1 + 2(p-1)$ be such that $\text{ER}(s)$ holds, and suppose that we have a polynomial $Q \in \text{Poly}_{\leq s+1}(V \rightarrow {}^*\mathbb{T})$ that is a function of the factor P , thus $Q = F(P_{1,1}, P_{2,0})$ for some $F : (\frac{1}{p^2}\mathbb{Z}/\mathbb{Z}) \times (\frac{1}{p}\mathbb{Z}/\mathbb{Z}) \rightarrow {}^*\mathbb{T}$. Then we can find a polynomial $R \in \text{Poly}_{\leq s+p}(V \rightarrow {}^*\mathbb{T})$ that is a root of Q (thus $pR = Q$), which is a function of P' (thus $R = F'(P_{1,2}, P_{2,1})$ for some $F' : (\frac{1}{p^3}\mathbb{Z}/\mathbb{Z}) \times (\frac{1}{p^2}\mathbb{Z} \times \mathbb{Z}) \rightarrow {}^*\mathbb{T}$). If s was equal to $D_1 + p - 2$, then R would be a linear combination of $P_{1,2}, P_{2,1}$ plus a function $F''(P_{1,1}, P_{2,0})$ of the lower degree polynomials $P_{1,2}, P_{2,1}$.

One consequence of the exact roots property is that it allows for a regularity lemma:

Lemma 9.6 (Regularity lemma). *Let $s_0 \geq 1$ be a standard integer such that $\text{ER}(s)$ and $\text{GIP}(s)$ hold for all $0 \leq s < s_0$. Suppose that $P = (P_1, \dots, P_m)$ are a bounded tuple of polynomials $P_1, \dots, P_m \in \text{Poly}_{\leq s_0}(V \rightarrow {}^*\mathbb{T})$ on a limit finite-dimensional vector space V . Then there exists a bounded index subspace V' of V and a regular factor $Q = (Q_{i,j})_{1 \leq i \leq m; 0 \leq j \leq J_i}$ on V' of degree at most s_0 such that all the P_i are measurable with respect to Q on V' . Furthermore, on V' , if P_i has degree d_i , then P_i is a linear combination (over \mathbb{Z}) of those polynomials $Q_{i,j}$ with $D_i + j(p-1) = d_i$, plus a function of the depth retraction $Q_{\leq d_i-1}$ of Q .*

Proof. We induct on s_0 . The claim is trivial for $s_0 \leq 1$ (note that any polynomial of degree 1 can be made constant by passing to a bounded index subspace), so suppose that $s_0 \geq 2$ and that the claim has already been proven for smaller s_0 . Observe from Lemma 1.7 that $p \text{Poly}_{\leq s_0}(V \rightarrow {}^*\mathbb{T}) \subset \text{Poly}_{\leq s_0, \text{BR}}(V \rightarrow {}^*\mathbb{T})$, thus the abelian group $\text{Poly}_{\leq s_0}(V \rightarrow {}^*\mathbb{T}) / \text{Poly}_{\leq s_0, \text{BR}}(V \rightarrow {}^*\mathbb{T})$ is in fact a vector space. This vector space contains $\text{Poly}_{\leq s_0}(V \rightarrow \iota(\mathbb{F})) / \text{Poly}_{\leq s_0, \text{BR}}(V \rightarrow \iota(\mathbb{F}))$ as a subspace.

As every finitely generated vector space has a finite basis, we may thus represent P_1, \dots, P_m as a linear combination (over \mathbb{Z}) of a bounded number of polynomials $P'_1, \dots, P'_{m'} \in \text{Poly}_{\leq s_0}(V \rightarrow {}^*\mathbb{T})$ that are linearly independent modulo $\text{Poly}_{\leq s_0, \text{BR}}(V \rightarrow {}^*\mathbb{T}) + \text{Poly}_{\leq s_0}(V \rightarrow \iota(\mathbb{F}))$, a bounded number of classical polynomials $R_1, \dots, R_{m''} \in \text{Poly}_{\leq s_0}(V \rightarrow \iota(\mathbb{F}))$ that are linearly independent modulo $\text{Poly}_{\leq s_0, \text{BR}}(V \rightarrow \iota(\mathbb{F}))$, and a bounded number of bounded rank polynomials $S_1, \dots, S_{m'''} \in \text{Poly}_{\leq s_0, \text{BR}}(V \rightarrow {}^*\mathbb{T})$. By definition of $\text{Poly}_{\leq s_0, \text{BR}}(V \rightarrow {}^*\mathbb{T})$, the $S_1, \dots, S_{m'''}$ are in turn functions of a bounded number of polynomials $S'_1, \dots, S'_{m''''} \in \text{Poly}_{\leq s_0-1}(V \rightarrow {}^*\mathbb{T})$.

By Lemma 1.7, the polynomials $pP'_1, \dots, pP'_{m'}$ have degree at most $\max(s_0 - p + 1, 0)$, and in particular have degree $\leq s_0 - 1$. Applying the induction hypothesis, and passing to a bounded index subspace V' of V , the polynomials $S'_1, \dots, S'_{m''''}, pP'_1, \dots, pP'_{m'}$, when restricted to V' , are then all functions of a single regular factor $Q = (Q_{i,j})_{1 \leq i \leq m''''; 0 \leq j \leq J_i}$ of degree $\leq s_0 - 1$ with initial degrees $D_1, \dots, D_{m''''} \geq 2$.

Henceforth all polynomials will be understood to be restricted to V' . Using Lemma 4.7, we see that the various linear independence properties on $P'_1, \dots, P'_{m'}$ and $R_1, \dots, R_{m''}$ descend from V to V' .

By using Lemma 1.7(v) to perform depth extensions on Q as necessary, we may assume that

$$s_0 - p < D_i + J_i(p - 1) \leq s_0 - 1$$

for all $i = 1, \dots, m$; of course, this keeps the degree of Q to be $\leq s_0 - 1$. Note that Q also remains regular, thanks to Lemma 9.3.

We now perform one further depth extension to Q to obtain $Q' := (Q_{i,j})_{1 \leq i \leq m''''; 0 \leq j \leq J_i+1}$ by choosing Q_{i,J_i+1} to be a polynomial in $\text{Poly}_{\leq D_i + (J_i+1)(p-1)}(V' \rightarrow {}^*\mathbb{T})$ obeying $pQ_{i,J_i+1} = Q_{i,J_i}$; such a polynomial is available thanks to Lemma 1.7(v). This is clearly a factor. By Lemma 9.3, Q' is also regular.

Note that when constructing Q' , we have the freedom to modify each Q_{i,J_i+1} additively by a classical polynomial from $\text{Poly}_{\leq D_i + (J_i+1)(p-1)}(V' \rightarrow \iota(\mathbb{F}))$; this freedom will be important later on.

Now consider the polynomials $pP'_1, \dots, pP'_{m'}$. By Lemma 1.7, they have degree $\leq \max(s_0 - p + 1, 0)$, and are also functions of Q . Applying the hypothesis $\text{ER}(s_0 - p)$ (and refining V' if necessary), we conclude that we can find polynomials $U_1, \dots, U_{m'} \in \text{Poly}_{\leq s_0}(V' \rightarrow {}^*\mathbb{T})$ that are functions of Q' and such that

$$pP'_l = pU_l$$

for all $1 \leq l \leq m'$. In other words, P'_l and U_l differ by a classical polynomial in $\text{Poly}_{\leq s_0}(V' \rightarrow \iota(\mathbb{F}))$. (Note that these claims are trivial when $s_0 \leq p - 1$; the hypothesis $\text{ER}(s_0 - p)$ is only needed when $s_0 > p - 1$). Furthermore, each U_l is a linear combination of those Q_{i,J_i+1} with $D_i + (J_i + 1)(p - 1) = s_0$, plus a function of $Q'_{\leq s_0-1}$, thus

$$U_l = \sum_{i \in A} c_{l,i} Q_{i,J_i+1} + E_l \tag{9.2}$$

for some coefficients $c_{l,i} \in \mathbb{Z}$, where E_l is a function of $Q_{\leq s_0-1}$, and A is the set of those $1 \leq i \leq m''''$ with $D_i + (J_i + 1)(p - 1) = s_0$. We claim that the vectors $\vec{c}_l := (c_{l,i} \bmod p)_{i \in A} \in \mathbb{F}^A$ for $l = 1, \dots, m'$ are linearly independent. Indeed, suppose for contradiction that we had a non-trivial linear dependence

$$a_1 \vec{c}_1 + \dots + a_{m'} \vec{c}_{m'} = 0$$

in \mathbb{F}^A for some $a_1, \dots, a_{m'} \in \{0, \dots, p - 1\}$, not all zero. Then by (9.2), the polynomial

$$a_1 U_1 + \dots + a_{m'} U_{m'}$$

is a function of $Q_{\leq s_0-1}$, and of degree $\leq s_0$. As the U_l differ from P'_l by an element of $\text{Poly}_{\leq s_0}(V' \rightarrow \iota(\mathbb{F}))$, we conclude that the $P'_1, \dots, P'_{m'}$ are linearly dependent modulo $\text{Poly}_{\leq s_0, \text{BR}}(V' \rightarrow {}^*\mathbb{T}) + \text{Poly}_{\leq s_0}(V' \rightarrow \iota(\mathbb{F}))$, contradicting the construction of $P'_1, \dots, P'_{m'}$. Thus the $\vec{c}_1, \dots, \vec{c}_{m'}$ are linearly independent.

We would like to modify Q' so that the U_l will agree with P'_l exactly. To do this, recall that we had the freedom to modify each of the Q_{i, J_i+1} by an arbitrary classical polynomial in $\text{Poly}_{\leq s_0}(V' \rightarrow \iota(\mathbb{F}))$; this modifies the U_l by a corresponding classical polynomial in $\text{Poly}_{\leq s_0}(V' \rightarrow \iota(\mathbb{F}))$. Because the $\vec{c}_1, \dots, \vec{c}_{m'}$ are linearly independent, the U_l can be so modified independently. Since P'_l and U_l already only differed by such a classical polynomial, we can thus modify each of the Q_{i, J_i+1} so that the U_l are equal to P'_l simultaneously for all $1 \leq i \leq m'$. Having done so, we now see that the P'_l are functions of $Q'_{\leq s_0}$.

Now we extend $Q'_{\leq s_0}$ to a further factor Q'' by adjoining the classical polynomials $R_1, \dots, R_{m''}$ as new dimensions of degree s_0 and depth 0. This is still a factor; we claim that it remains regular. To see this, we need to show that the $R_1, \dots, R_{m''}$, together with the $Q_{i, j}$ with $D_i + j(p-1) = s_0$, are linearly independent modulo $\text{Poly}_{\leq s_0, \text{BR}}(V' \rightarrow {}^*\mathbb{T})$. Note from construction that all such $Q_{i, j}$ must be of the form Q_{i, J_i+1} . Suppose for contradiction that there was a non-trivial linear dependence. As the $R_1, \dots, R_{m''}$ were already linearly independent modulo $\text{Poly}_{\leq s_0, \text{BR}}(V \rightarrow {}^*\mathbb{T})$, this dependence must involve at least one of the Q_{i, J_i+1} . Multiplying by p and using Corollary 5.2, we see that there is a dependence among those Q_{i, J_i} with $D_i + (J_i + 1)(p-1) = s_0$ modulo $\text{Poly}_{\leq s_0-p+1, \text{BR}}(V' \rightarrow {}^*\mathbb{T})$, but this contradicts the regularity of Q .

By construction, all the polynomials $P'_1, \dots, P'_{m'}, R_1, \dots, R_{m''}, S_1, \dots, S_{m'''}$ are functions of the regular factor Q'' , which has degree $\leq s_0$, and so P_1, \dots, P_m are functions of Q'' also. A careful inspection of the above argument also shows that each P_i was in fact an integer linear combination of those $Q''_{i, j}$ in Q'' with $D_i + j(p-1) = d_i$, plus a function of $Q''_{\leq d_i-1}$ (the cases $d_i = s_0$ and $d_i < s_0$ have to be treated separately). The claim follows. \square

In view of the regularity lemma, Theorem 4.2 now follows from

Theorem 9.7 (Exact roots, technical version). *Let $s_0 \geq 1$ be such that $\text{ER}(s_0 - 1)$ holds, and $\text{GIP}(s)$ holds for all $s \leq s_0$. Then $\text{ER}(s_0)$ holds.*

Indeed, assuming Theorem 9.7, then in the situation in Theorem 4.2, we have $\text{ER}(s')$ for all $0 \leq s' \leq s$ by strong induction (the case $\text{ER}(0)$ being trivial). If $P \in \text{Poly}_{\leq s'+1, \text{BR}}(V \rightarrow {}^*\mathbb{T})$ for some $0 \leq s' \leq s$, then by Lemma 9.6, P is a function of a regular factor R of degree $\leq s'$. We then create a depth extension R' of R by extending all the depths J'_i so that $s' < D_i + J'_i(p-1) \leq s' + p - 1$; the existence of such a depth extension is guaranteed by Lemma 1.7(v). Applying $\text{ER}(s)$, we conclude that we can find $Q \in \text{Poly}_{\leq s'+p}(V \rightarrow {}^*\mathbb{T})$ that is a function of R' such that $pQ = P$. Since R' has degree $\leq s' + p - 1$, Q has bounded rank, and Theorem 4.2 follows.

It remains to establish Theorem 9.7. To do this, we first require an equidistribution lemma, analogous to (but more complicated than) Lemma 7.8.

Let $P = (P_{i,j})_{1 \leq i \leq m; 0 \leq j \leq J_i}$ be a factor with initial degrees D_1, \dots, D_m . By (9.1), the polynomials in P are in fact functions of $\tilde{P} := (P_{i,J_i})_{1 \leq i \leq m}$, which we interpret as a map from V to ${}^*\mathbb{T}^m$.

It is now convenient to use the machinery of polynomial algebra, which is reviewed in Section C. We define a filtration $({}^*\mathbb{T}^m)_{\mathbb{N}}$ on ${}^*\mathbb{T}^m$ by defining $({}^*\mathbb{T}^m)_k$ for each standard natural number k to be the (finite) group generated by the elements $\frac{1}{p^{J_i-j+1}}e_i \bmod {}^*\mathbb{Z}^m$ for which $1 \leq i \leq m$, $0 \leq j \leq J_i$, and $k \leq D_i + j(p-1)$, where e_1, \dots, e_m is the standard basis of \mathbb{R}^m . This is easily seen to be a filtration (see Example C.3). We call this the *filtration with depths J_1, \dots, J_m and initial degrees D_1, \dots, D_m* .

Lemma 9.8 (Equidistribution of regular factors). *Suppose that $\text{GIP}(s)$ is true for all $0 \leq s \leq s_0$, let $P = (P_{i,j})_{1 \leq i \leq m; 0 \leq j \leq J_i}$ be a factor of degree $\leq s_0$, and let $k \geq 0$ be a standard integer. Set $\tilde{P} := (P_{i,J_i})_{1 \leq i \leq m}$, and let ${}^*\mathbb{T}^m$ be given the filtration with depths J_1, \dots, J_m and initial degrees D_1, \dots, D_m . We give V the maximal degree ≤ 1 filtration (see Example C.3).*

- (i) $\tilde{P} : V \rightarrow {}^*\mathbb{T}^m$ is a polynomial map (see Definition C.6).
- (ii) If furthermore P is regular, then $\tilde{P} : V \rightarrow {}^*\mathbb{T}^m$ is equidistributed on cubes (see Definition C.21).

Proof. To prove (i), it suffices by Definition C.6 to verify that

$$\Delta_{h_1} \dots \Delta_{h_k} \tilde{P}(x) \in ({}^*\mathbb{T}^m)_k \quad (9.3)$$

whenever $k \in \mathbb{N}$ and $h_1, \dots, h_k, x \in V$. But if $1 \leq i \leq m$, then we clearly have

$$\Delta_{h_1} \dots \Delta_{h_k} P_{i,J_i}(x) \in \frac{1}{p^{J_i+1}}\mathbb{Z}/\mathbb{Z}$$

since P_{i,J_i} takes values in $\frac{1}{p^{J_i+1}}\mathbb{Z}/\mathbb{Z}$; and if $D_i + j(p-1) < k \leq D_i + (j+1)(p-1)$ for some $0 \leq j \leq J_i$ then

$$\Delta_{h_1} \dots \Delta_{h_k} P_{i,j}(x) = 0$$

and since $p^{J_i-j}P_{i,J_i} = P_{i,j}$ we get

$$\Delta_{h_1} \dots \Delta_{h_k} P_{i,J_i}(x) \in \frac{1}{p^{J_i-j}}\mathbb{Z}/\mathbb{Z}.$$

Comparing this with the definition of $({}^*\mathbb{T}^m)_k$ we obtain (9.3).

Now we verify (ii), which is trickier. We need to show that the map $\text{HK}^k(\tilde{P}) : \text{HK}^k(V) \rightarrow \text{HK}^k({}^*\mathbb{T}^m)$ is equidistributed. By the Weyl equidistribution criterion (Lemma C.23), it suffices to show that

$$\mathbb{E}_{x, h_1, \dots, h_k \in V} e(\eta(\text{HK}^k(\tilde{P})(x, h_1, \dots, h_k))) = o(1) \quad (9.4)$$

whenever $\eta : \text{HK}^k({}^*\mathbb{T}^m) \rightarrow \mathbb{T}$ is a non-zero homomorphism.

Observe that $\mathrm{HK}^k(*\mathbb{T}^m)$ is a (finite) subgroup of $(\mathbb{T}^m)^{\{0,1\}^k}$. By Pontryagin duality, η must therefore be the restriction of a homomorphism from $(\mathbb{T}^m)^{\{0,1\}^k}$ to \mathbb{T} . In other words, we can find (non-unique) integers $c_{i,\omega}$ for $1 \leq i \leq m$ and $\omega \in \{0,1\}^m$ such that

$$\eta((t_{i,\omega})_{1 \leq i \leq m; \omega \in \{0,1\}^k}) = \sum_{i=1}^m \sum_{\omega \in \{0,1\}^k} c_{i,\omega} t_{i,\omega} \quad (9.5)$$

for all $(t_{i,\omega})_{1 \leq i \leq m; \omega \in \{0,1\}^k}$ in $\mathrm{HK}^k(*\mathbb{T}^m)$. In particular, the left-hand side of (9.4) becomes

$$\mathbb{E}_{x, h_1, \dots, h_k \in V} e\left(\sum_{i=1}^m \sum_{\omega \in \{0,1\}^k} c_{i,\omega} P_{i,J_i}(x + \omega_1 h_1 + \dots + \omega_k h_k)\right). \quad (9.6)$$

The coordinates $t_{i,\omega} \in *\mathbb{T}$ of points $(t_{i,\omega})_{1 \leq i \leq m; \omega \in \{0,1\}^k}$ in $\mathrm{HK}^k(*\mathbb{T}^m)$ obey a number of constraints. Firstly, for each ω , $(t_{i,\omega})_{1 \leq i \leq m}$ must lie in $(*\mathbb{T}^m)_0$, or in other words we have

$$p^{J_i+1} t_{i,\omega} = 0 \quad (9.7)$$

whenever $1 \leq i \leq m$ and $\omega \in \{0,1\}^k$. Secondly, from Proposition C.16 we see that

$$p^{J_i-j} \sum_{\omega \in F} (-1)^{|\omega|} t_{i,\omega} = 0 \quad (9.8)$$

whenever $1 \leq i \leq m$, $\omega \in \{0,1\}^k$, $0 \leq j \leq J_i$, and F is a face in $\{0,1\}^k$ of dimension greater than $D_i + j(p-1)$. (In fact, Proposition C.16 asserts that these are the *only* constraints on the $t_{i,\omega}$.)

We can use these constraints to place the coefficients $c_{i,\omega}$ in a “reduced form”, as follows. First observe from (9.8) that if there exist $1 \leq i \leq m$, $\omega \in \{0,1\}^k$, and $0 \leq j \leq J_i$ with $|\omega| > D_i + j(p-1)$ and $c_{i,\omega} \geq p^{J_i-j}$ or $c_{i,\omega} < 0$ then by adding a suitable multiple of (9.8) to (9.5), one can place $c_{i,\omega}$ in the interval $\{0, \dots, p^{J_i-j} - 1\}$, at the expense of changing the values of $c_{i,\omega'}$ for various ω' with $|\omega'| < |\omega|$. Iterating this procedure (starting with those ω with large values of $|\omega|$ and then working downward) we may assume without loss of generality that

$$0 \leq c_{i,\omega} < p^{J_i-j} \quad (9.9)$$

whenever $1 \leq i \leq m$, $\omega \in \{0,1\}^k$, and $0 \leq j \leq J_i$ is such that $|\omega| > D_i + j(p-1)$.

In a similar spirit, by using (9.7), we may assume that

$$0 \leq c_{i,\omega} < p^{J_i+1}$$

for all $1 \leq i \leq m$ and $\omega \in \{0,1\}^k$.

Since η is non-zero, at least one of the $c_{i,\omega}$ is non-zero. Let $\omega_* \in \{0,1\}^k$ be such that c_{i,ω_*} is non-zero for at least one $1 \leq i \leq m$, and such that $|\omega_*|$ is maximal with respect to this property. By permutation symmetry we may assume that $\omega_* = 1^K 0^{k-K} = (1, \dots, 1, 0, \dots, 0)$ for some $0 \leq K \leq k$. The expression (9.6) can then be factored as

$$\mathbb{E}_{x, h_1, \dots, h_k \in V} e(Q(x + h_1 + \dots + h_K)) \prod_{l=1}^K f_l(x, h_1, \dots, h_k)$$

where $Q : V \rightarrow {}^*\mathbb{T}$ is the expression

$$Q := \sum_{i=1}^m c_{i,\omega_*} P_{i,J_i}$$

and each $f_l : V^k \rightarrow {}^*\mathbb{C}$ is a limit function bounded in magnitude by 1 and independent of h_l . Using the second Cauchy-Schwarz-Gowers inequality (Lemma B.1(v)), it thus suffices to show that

$$\|e(Q)\|_{U^K(V)} = o(1).$$

Let $I := \{1 \leq i \leq m : c_{i,\omega_*} \neq 0\}$, then I is non-empty. For each $i \in I$, let $0 \leq j_i \leq J_i$ be the least integer j such that $p^{J_i-j} | c_{i,\omega_*}$. Since $p^{J_i-j_i} P_{i,J_i} = P_{i,j_i}$, we thus have

$$Q = \sum_{i \in I} a_i P_{i,j_i} \tag{9.10}$$

for some integers a_i that are not divisible by p .

Let

$$D := \sup_{i \in I} D_i + j_i(p-1).$$

As P has degree $\leq s_0$, we have $D \leq s_0$. Also, since each P_{i,j_i} has degree $D_i + j_i(p-1)$, we see that Q has degree $\leq D$.

If one had

$$D_i + j_i(p-1) < |\omega_*| = K$$

for some $i \in I$, then from (9.9) we would have $c_{i,\omega_*} < p^{J_i-j_i}$, a contradiction; so we must have $D_i + j_i(p-1) \geq K$ for all $i \in I$. In particular, $D \geq K$. By the monotonicity of the Gowers norms (Lemma B.1(ii)), it thus suffices to show that

$$\|e(Q)\|_{U^D(V)} = o(1).$$

Applying the induction hypothesis $\text{GIP}(D-1)$, it thus suffices to show that $Q \notin \text{Poly}_{\leq D, \text{BR}}(V \rightarrow {}^*\mathbb{T})$. But, as P is regular, the polynomials P_{i,j_i} with $D_i + j_i(p-1) = D$ are linearly independent over \mathbb{F} modulo $\text{Poly}_{\leq D, \text{BR}}(V \rightarrow {}^*\mathbb{T})$, and the polynomials P_{i,j_i} with $D_i + j_i(p-1) < D$ already lie in $\text{Poly}_{\leq D, \text{BR}}(V \rightarrow {}^*\mathbb{T})$. Since there is at least $i \in I$ with $D_i + j_i(p-1) = D$, and all coefficients a_i in (9.10) are nonzero modulo p , we obtain the desired claim. \square

From the above corollary and Corollary C.22 we conclude

Corollary 9.9 (Polynomials on a regular factor). *Suppose that $\text{GIP}(s)$ is true for all $0 \leq s \leq s_0$, let $P = (P_{i,j})_{1 \leq i \leq m; 0 \leq j \leq J_i}$ be a regular factor of degree $\leq s_0$, let $\tilde{P} : V \rightarrow {}^*\mathbb{T}^m$ be the associated map $\tilde{P} = (P_{i,J_i})_{1 \leq i \leq m}$, let $d \geq 0$ be an integer, and let $f : {}^*\mathbb{T}^m \rightarrow {}^*\mathbb{T}$ be a function. Then the following are equivalent:*

- $f(\tilde{P}) : V \rightarrow {}^*\mathbb{T}$ is a polynomial of degree $\leq d$.
- $f : {}^*\mathbb{T}^m \rightarrow {}^*\mathbb{T}$ is a polynomial map (where we give ${}^*\mathbb{T}^m$ the filtration with depths J_1, \dots, J_m and initial degrees D_1, \dots, D_m , and ${}^*\mathbb{T}$ the maximal degree $\leq d$ filtration, see Example C.3).

We now lift the conclusion of Corollary 9.9 from ${}^*\mathbb{T}^m$ to \mathbb{Z}^m . Given natural numbers D_1, \dots, D_m , we define the *filtration* $(\mathbb{Z}^m)_{\mathbb{N}}$ of *initial degrees* D_1, \dots, D_m on \mathbb{Z}^m by setting $(\mathbb{Z}^m)_k$, for each natural number k , to be the subgroup of \mathbb{Z}^m generated by those elements $p^j e_i$ for which $1 \leq i \leq m$, $j \in \mathbb{N}$, and $k \leq D_i + j(p-1)$; this is easily seen to be a filtration.

Corollary 9.10 (Polynomials on a regular factor, again). *Suppose that $\text{GIP}(s)$ is true for all $0 \leq s \leq s_0$, let $P = (P_{i,j})_{1 \leq i \leq m; 0 \leq j \leq J_i}$ be a regular factor of degree $\leq s_0$ and initial degrees D_1, \dots, D_m , let $d \geq 0$ be an integer. Let $Q : V \rightarrow {}^*\mathbb{T}$ be a limit function. Then the following are equivalent:*

- Q is measurable with respect to P , and is a polynomial of degree $\leq d$.
- There exists a polynomial map $\tilde{f} : \mathbb{Z}^m \rightarrow {}^*\mathbb{T}$ from \mathbb{Z}^m (with the filtration of initial degrees D_1, \dots, D_m) to ${}^*\mathbb{T}$ (with the maximal degree $\leq d$ filtration), such that \tilde{f} is periodic with period $p^{J_i+1} e_i$ for each $1 \leq i \leq m$, and such that one has

$$Q(x) = \tilde{f}(a_1, \dots, a_m), \quad (9.11)$$

whenever $x \in V$ and $a_1, \dots, a_m \in \mathbb{Z}$ are such that

$$P_{i,J_i}(x) = \frac{a_i}{p^{J_i+1}} \bmod 1, \quad (9.12)$$

for all $1 \leq i \leq m$.

Proof. If Q is measurable with respect to P and is a polynomial of degree $\leq d$, then by Corollary 9.9, we can write $Q = f(\tilde{P})$, where $f : {}^*\mathbb{T}^m \rightarrow {}^*\mathbb{T}$ is a polynomial map from ${}^*\mathbb{T}^m$ (with the filtration of initial degrees D_1, \dots, D_m and depths J_1, \dots, J_m) to ${}^*\mathbb{T}$ (with the maximal degree $\leq d$ filtration).

Let $\phi : \mathbb{Z}^m \rightarrow {}^*\mathbb{T}^m$ be the map

$$\phi(a_1, \dots, a_m) := \left(\frac{a_1}{p^{J_1+1}} \bmod 1, \dots, \frac{a_m}{p^{J_m+1}} \bmod 1 \right).$$

One easily verifies that this is a polynomial map (indeed, it is a filtered homomorphism) from \mathbb{Z}^m (with the filtration of initial degrees D_1, \dots, D_m) to ${}^*\mathbb{T}^m$ (with the filtration of initial degrees D_1, \dots, D_m and depths J_1, \dots, J_m). Thus the function $\tilde{f} : \mathbb{Z}^m \rightarrow {}^*\mathbb{T}$ defined by $\tilde{f} := f \circ \phi$ is also a polynomial map from \mathbb{Z}^m (with the filtration of initial degrees D_1, \dots, D_m) to ${}^*\mathbb{T}$ (with the maximal degree $\leq d$ filtration). It is also periodic with period $p^{J_i+1} e_i$ for each $1 \leq i \leq m$, because ϕ is also periodic with these periods. By construction one also has (9.11) whenever (9.12). This proves one implication of the corollary. The other implication follows by reversing the above argument (noting that ϕ is weakly equidistributed on cubes in the sense of Definition C.18, so that one can apply Lemma C.19). \square

In view of the above corollary, Theorem 9.7 can now be deduced from an analogous result on the integer lattice \mathbb{Z}^m , which we formulate precisely as follows:

Proposition 9.11 (Exact roots in \mathbb{Z}^m). *Let $m \geq 0$, $D_1, \dots, D_m \geq 1$, and $d \geq 0$ be standard natural numbers. Let $\tilde{f} : \mathbb{Z}^m \rightarrow {}^*\mathbb{T}$ be a polynomial map from \mathbb{Z}^m (with the*

filtration of initial degrees D_1, \dots, D_m) to ${}^*\mathbb{T}$ (with the maximal degree $\leq d$ filtration). Then we can find a polynomial map $\tilde{g} : \mathbb{Z}^m \rightarrow {}^*\mathbb{T}$ from \mathbb{Z}^m (with the filtration of initial degrees D_1, \dots, D_m) to \mathbb{T} (with the maximal degree $\leq d + p - 1$ filtration) such that $p\tilde{g} = \tilde{f}$. Furthermore, \tilde{g} is a linear combination (over \mathbb{Z}) of the functions $(a_1, \dots, a_m) \mapsto \frac{a_i}{p^{j+1}} \bmod 1$ with $1 \leq i \leq m$ and $D_i + j(p-1) = d+p-1$, plus a function of the expressions $a_i \bmod p^{j+1}$ with $1 \leq i \leq m$ and $D_i + j(p-1) < d+p-1$. In particular, \tilde{g} is periodic with period $p^{j+1}e_i$ whenever $1 \leq i \leq m$ and $D_i + j(p-1) > d$.

We remark that in contrast with the previous arguments, the above proposition holds even when the D_i are equal to 1 (indeed, Lemma 1.7(iii) can be viewed as a special case in which $D_1 = \dots = D_m = 1$ and the functions are periodic with period $p\mathbb{Z}^m$).

We will prove this Proposition in the next section. For now, we show how Proposition 9.11 implies Theorem 9.7 and thus Theorem 4.2.

Proof of Theorem 9.7 assuming Proposition 9.11. Let $s_0 \geq 1$ be such that $\text{ER}(s_0 - 1)$ holds, and $\text{GIP}(s)$ holds for all $s \leq s_0$. Let P be a regular factor of degree at most s_0 and some initial degrees D_1, \dots, D_m , let $Q \in \text{Poly}_{\leq s_0+1}(V \rightarrow {}^*\mathbb{T})$ be a function of P , and let $P' = (P_{i,j})_{1 \leq i \leq m; 0 \leq j \leq J'_i}$ be a depth extension of P with $D_i + J'_i(p-1) > s_0$ for all $1 \leq i \leq m$. Our objective is to find a polynomial $R \in \text{Poly}_{\leq s+p}(V \rightarrow {}^*\mathbb{T})$ that is a function of P' such that $pR = Q$.

Note from Lemma 9.3 that P' is automatically regular.

By Corollary 9.10, we can find a function $\tilde{f} : \mathbb{Z}^m \rightarrow {}^*\mathbb{T}$ of weighted degree $\leq s_0 + 1$ which is periodic with period $p^{J_i+1}e_i$ for each $i = 1, \dots, m$, such that one has

$$Q(x) = \tilde{f}(a_1, \dots, a_m)$$

whenever $x \in V$ and $a_1, \dots, a_m \in \mathbb{Z}$ are such that

$$P_{i,J_i}(x) = \frac{a_i}{p^{J_i+1}} \bmod 1$$

for all $1 \leq i \leq m$.

Applying Proposition 9.11, one can find a function $\tilde{g} : \mathbb{Z}^m \rightarrow {}^*\mathbb{T}$ of weighted degree $\leq s_0 + p$ which is periodic with period $p^{j+1}e_i$ whenever $1 \leq i \leq m$ and $D_i + j(p-1) > s_0 + 1$, such that $p\tilde{g} = \tilde{f}$. In particular, \tilde{g} is periodic with period $p^{J'_i+1}e_i$ for each $1 \leq i \leq m$. We may therefore define the function $R : V \rightarrow {}^*\mathbb{T}$ by setting

$$R(x) := \tilde{g}(a_1, \dots, a_m)$$

whenever $x \in V$ and $a_1, \dots, a_m \in \mathbb{Z}$ are such that

$$P_{i,J'_i}(x) = \frac{a_i}{p^{J'_i+1}} \bmod 1$$

for all $1 \leq i \leq m$; the periodicity properties of \tilde{g} ensure that R is well-defined. By Corollary 9.10, R is a function of P' which is a polynomial of degree $\leq s_0 + p$. By construction, one has $pR = Q$, and the claim follows. \square

The only remaining task is to establish Proposition 9.11. This will be the subject of the next section.

10. EXACT ROOTS FOR POLYNOMIALS ON \mathbb{Z}^m

We now prove Proposition 9.11. Throughout this section the dimension $m \geq 0$ and initial degrees $D_1, \dots, D_m \geq 1$ are fixed.

It is convenient to rephrase the polynomiality condition in terms of derivatives. Define a *basic generator* v to be an element of \mathbb{Z}^m of the form $p^j e_i$, where $1 \leq i \leq m$ and $j \geq 0$. Define a *multigenerator* to be a tuple $\vec{v} = (v_1, \dots, v_r)$ of basic generators, where $r \geq 0$ is a standard natural number. We associate to each basic generator $p^j e_i$ a *weighted degree* $\deg(p^j e_i) := D_i + j(p-1)$, and associate to each multigenerator $\vec{v} = (v_1, \dots, v_r)$ a weighted degree $\deg(\vec{v}) := \deg(v_1) + \dots + \deg(v_r)$. We also associate to \vec{v} the differential operator

$$\Delta_{\vec{v}} := \Delta_{v_1} \dots \Delta_{v_r}.$$

We say that a function $\tilde{f} : \mathbb{Z}^m \rightarrow {}^*\mathbb{T}$ has *weighted degree* $\leq d$ if one has $\Delta_{\vec{v}} \tilde{f} = 0$ whenever $\deg(\vec{v}) > d$. In other words, we have

$$\left(\prod_{i=1}^m \prod_{j=0}^{\infty} \Delta_{p^j e_i}^{a_{i,j}} \right) \tilde{f} = 0$$

whenever $a_{i,j}$ are natural numbers (at most finitely many of which are non-zero) with $\sum_{i=1}^m \sum_{j=0}^{\infty} a_{i,j} (D_i + j(p-1)) > d$.

From Proposition C.10 we have

Proposition 10.1 (Differential characterisation of polynomiality). *Let $\tilde{f} : \mathbb{Z}^m \rightarrow {}^*\mathbb{T}$ be a function, and let $d \geq 0$ be a standard natural number. Then the following are equivalent:*

- (i) \tilde{f} is a polynomial map from \mathbb{Z}^m (with the filtration of initial degrees D_1, \dots, D_m) to ${}^*\mathbb{T}$ (with the maximal degree $\leq d$ filtration).
- (ii) \tilde{f} has weighted degree $\leq d$.

One nice feature of a weighted bounded degree polynomials is that they have some periodicity properties:

Lemma 10.2 (Periodicity properties). *Let $\tilde{f} : \mathbb{Z}^m \rightarrow {}^*\mathbb{T}$ be of weighted degree $\leq d$.*

- (i) *For any $1 \leq i \leq m$, \tilde{f} is periodic with period $p^j e_i$ whenever $1 \leq i \leq m$ and $j \geq 0$ are such that $D_i + j(p-1) > d$.*
- (ii) *\tilde{f} is a linear combination (over \mathbb{Z}) of the functions $(a_1, \dots, a_m) \mapsto \frac{a_i}{p^{j+1}} \bmod 1$ with $1 \leq i \leq m$ and $D_i + j(p-1) = d$, plus a function of the expressions $a_i \bmod p^{j+1}$ with $1 \leq i \leq m$ and $D_i + j(p-1) < d$.*

Proof. If $D_i + j(p-1) > d$ then $\Delta_{p^j e_i} \tilde{f} = 0$ by Proposition 10.1. This proves (i).

Now we turn to (ii). Let I denote the set of all $1 \leq i \leq m$ for which there is a natural number j_i for which $D_i + j_i(p-1) = d$. If $i \in I$, then $\Delta_{p^{j_i} e_i} \tilde{f}$ has weighted degree ≤ 0 and is thus constant. On the other hand, by (i), \tilde{f} is periodic with period $p^{j_i+1} e_i$. We conclude that $\Delta_{p^{j_i} e_i} \tilde{f} = c_i/p$ for some $c_i \in \{0, \dots, p-1\}$. Write $\tilde{g} := \sum_{i \in I} c_i \frac{a_i}{p^{j_i+1}} \bmod 1$, then we see that \tilde{g} is periodic with period $p^{j_i} e_i$ for each $i \in I$, and thus (by (i)), is also periodic with period $p^j e_i$ whenever $1 \leq i \leq m$ and $D_i + j(p-1) < d$. \square

From Proposition 10.1 and Lemma 10.2 we may thus rephrase Proposition 9.11 as follows:

Proposition 10.3 (Exact roots in \mathbb{Z}^m , again). *Let $d \geq 0$ be standard. Let $\tilde{f} : \mathbb{Z}^m \rightarrow {}^*\mathbb{T}$ be a map of weighted degree $\leq d$. Then we can find a map $\tilde{g} : \mathbb{Z}^m \rightarrow {}^*\mathbb{T}$ of weighted degree $\leq d + p - 1$ such that $p\tilde{g} = \tilde{f}$.*

In order to prove Proposition 10.3, we use the following explicit description of those functions $\tilde{f} : \mathbb{Z}^m \rightarrow {}^*\mathbb{T}$ of a given weighted degree, which generalises Lemma 1.7(iii):

Proposition 10.4 (Classification of polynomials). *Let $d \geq 0$ be standard, and let $\tilde{f} : \mathbb{Z}^m \rightarrow {}^*\mathbb{T}$ be a map. Then the following are equivalent:*

- (i) \tilde{f} has weighted degree $\leq d$.
- (ii) \tilde{f} can be expressed as

$$\tilde{f}(x_1, \dots, x_m) = \alpha + \sum_{\substack{i_1, \dots, i_m \geq 0; r \geq 0 : \\ (\sum_{j=1}^m D_j i_j) + r(p-1) \leq d}} \frac{c_{i_1, \dots, i_m, r}}{p^{r+1}} \binom{x_1}{i_1} \dots \binom{x_m}{i_m} \bmod 1$$

for some $\alpha \in {}^*\mathbb{T}$ and integers $c_{i_1, \dots, i_m, r}$.

Proposition 10.4 immediately implies Proposition 10.3, since any element $\alpha \in {}^*\mathbb{T}$ has a p^{th} root, and any “monomial” $\frac{c_{i_1, \dots, i_m, r}}{p^{r+1}} \binom{x_1}{i_1} \dots \binom{x_m}{i_m}$ of degree $\leq d$ has a p^{th} root $\frac{c_{i_1, \dots, i_m, r}}{p^{r+2}} \binom{x_1}{i_1} \dots \binom{x_m}{i_m}$ of degree $\leq d + p - 1$.

Proof. We first show that (ii) implies (i). As each constant function α clearly is of degree $\leq d$, it suffices by linearity to show that the multinomial

$$M : (x_1, \dots, x_m) \mapsto \frac{1}{p^{r+1}} \binom{x_1}{i_1} \dots \binom{x_m}{i_m} \bmod 1$$

has weighted degree $\leq (\sum_{j=1}^m D_j i_j) + r(p-1)$ for any $i_1, \dots, i_m \geq 0$ and $r \in \mathbb{Z}$. (For inductive reasons we include the case when r is negative, but the claim is trivial in those cases as the multinomial M then vanishes modulo 1.)

We prove this by induction on the weighted degree $d_M := (\sum_{j=1}^m D_j i_j) + r(p-1)$. When $d_M \leq 0$ or less the claim is trivial, so suppose that d_M is positive, and that the claim has already been proven for smaller values of d_M .

To show that M has weighted degree $\leq d_M$, it then suffices to show that $\Delta_{p^k e_j} M$ has weighted degree $\leq d_M - D_j - k(p-1)$ for each $1 \leq j \leq m$ and $k \geq 0$. For sake of argument we shall just verify this when $j = m$, though the other cases are of course similar. From the binomial identity (6.5) one has

$$\Delta_{p^k e_m} M(x_1, \dots, x_m) = \sum_{l=1}^{i_m} \frac{1}{p^{r+1}} \binom{x_1}{i_1} \cdots \binom{x_m}{i_m - l} \binom{p^k}{l} \pmod{1},$$

so it will suffice to show that each term

$$\frac{1}{p^{r+1}} \binom{x_1}{i_1} \cdots \binom{x_m}{i_m - l} \binom{p^k}{l} \pmod{1} \quad (10.1)$$

with $1 \leq l \leq i_m$ has weighted degree $\leq d_M - D_m - k(p-1)$.

Fix l . We may assume that $l \leq p^k$, since the binomial coefficient $\binom{p^k}{l}$ vanishes otherwise. Let t be the largest natural number such that p^t divides l , then $t \leq k$. Inspecting the binomial coefficient

$$\binom{p^k}{l} = \frac{p^k}{l} \frac{p^k - 1}{1} \frac{p^k - 2}{2} \cdots \frac{p^k - l + 1}{l - 1}$$

we see that p^{k-t} divides $\binom{p^k}{l}$. Absorbing this factor into the $\frac{1}{p^{r+1}}$ term in (10.1) and using the induction hypothesis, we conclude that (10.1) has weighted degree

$$\leq \left(\sum_{j=1}^{m-1} D_j i_j \right) + D_m(i_m - l) + (r - k + t)(p-1).$$

But note that

$$D_m(l-1) \geq (l-1) \geq p^t - 1 \geq t(p-1),$$

and the claim follows.

Next, we show that (i) implies (ii). This claim is trivial for $d \leq 0$, so suppose inductively that $d > 0$ and that the claim has been proven for smaller values of d . We then fix d and assume as a second induction hypothesis that the claim has already been proven for smaller dimensions than m . We may assume that $m > 0$, since the $m = 0$ case is trivial.

Let \tilde{f} be of weighted degree $\leq d$, and consider the derivative $\Delta_{e_m} \tilde{f}$. By (i), this function has weighted degree $\leq d - D_m$, and thus by the induction hypothesis has a representation of the form

$$\Delta_{e_m} \tilde{f}(x_1, \dots, x_m) = \alpha + \sum_{\substack{i_1, \dots, i_m \geq 0; r \geq 0 : \\ (\sum_{j=1}^m D_j i_j) + r(p-1) \leq d - D_m}} \frac{c_{i_1, \dots, i_m, r}}{p^{r+1}} \binom{x_1}{i_1} \cdots \binom{x_m}{i_m} \pmod{1}.$$

We now introduce the function

$$\tilde{g}(x_1, \dots, x_m) := \sum_{\substack{i_1, \dots, i_m \geq 0; r \geq 0 : \\ (\sum_{j=1}^m D_j i_j) + r(p-1) \leq d - D_m}} \frac{c_{i_1, \dots, i_m, r}}{p^{r+1}} \binom{x_1}{i_1} \cdots \binom{x_{m-1}}{i_{m-1}} \binom{x_m}{i_m + 1} \pmod{1}.$$

As (ii) implies (i), we know that \tilde{g} has weighted degree $\leq d$. From Pascal's identity we have

$$\Delta_{e_m} \tilde{g}(x_1, \dots, x_m) = \sum_{\substack{i_1, \dots, i_m \geq 0; r \geq 0 : \\ (\sum_{j=1}^m D_j i_j) + r(p-1) \leq d - D_m}} \frac{c_{i_1, \dots, i_m, r}}{p^{r+1}} \binom{x_1}{i_1} \dots \binom{x_m}{i_m} \pmod{1}$$

and thus

$$\Delta_{e_m} \tilde{f} = \alpha + \Delta_{e_m} \tilde{g}.$$

We thus have

$$\tilde{f}(x_1, \dots, x_m) = \alpha x_m + \tilde{f}(x_1, \dots, x_{m-1}, 0) + \tilde{g}(x_1, \dots, x_m).$$

As \tilde{f} has weighted degree $\leq d$, the $m-1$ -dimensional function $(x_1, \dots, x_{m-1}) \mapsto \tilde{f}(x_1, \dots, x_{m-1}, 0)$ does also. By the second induction hypothesis, $\tilde{f}(x_1, \dots, x_{m-1}, 0)$ is already of the required form for (ii), while \tilde{g} is also of the required form by construction. It remains to show that the term αx_m has the required form.

By linearity, αx_m is of weighted degree $\leq d$. By Lemma 10.2, we thus see that $p^j \alpha = 0$ whenever $D_m + j(p-1) > d$. If we thus let j be the first natural number for which $D_m + j(p-1) > d$, then αx_m is a multiple of $\frac{1}{p^j} \binom{x_m}{1}$; as (ii) implies (i), this has degree $\leq D_m + (j-1)(p-1) \leq d$, and the claim follows. \square

The proof of Proposition 10.3 (and thus Proposition 9.11) is now complete.

11. DEDUCING THE INVERSE CONJECTURE FROM THE INVERSE CONJECTURE FOR POLYNOMIALS

In this section we deduce Theorem 1.11 from Theorem 1.20. This deduction can be done in either a finitary or an infinitary setting. In the finitary setting, one uses structural decomposition theorems as in [40], [13], [22], [15], [16]; the arguments in [16] are particularly close to those here. In the infinitary setting one can proceed by analogous decomposition theorems based on conditional expectation. We shall follow the latter approach here, in order to illustrate the parallel nature of the two arguments. (This latter approach is also adopted in [37].)

We first give a general abstract structural decomposition.

Lemma 11.1 (Decomposition). *Let V be a limit finite set, and let \mathcal{S} be a family of limit functions $P : V \rightarrow X$ on V , each of which takes only a finite number of values. (We do not assume that \mathcal{S} is itself a limit set.) Let $f : V \rightarrow {}^*\mathbb{C}$ be a limit function bounded in magnitude by some standard real A . Then one can decompose*

$$f = f_{\text{str}} + f_{\text{psd}}$$

where $f_{\text{str}}, f_{\text{psd}} : V \rightarrow {}^*\mathbb{C}$ are limit functions bounded in magnitude by A and $2A$ respectively, with the following properties:

- (f_{str} almost structured) For every standard $\varepsilon > 0$, one can find a function $f_\varepsilon : V \rightarrow {}^*\mathbb{C}$ that is a function of boundedly many functions from \mathcal{S} , such that f_ε is bounded in magnitude by A and $\|f_{\text{str}} - f_\varepsilon\|_{L^2(V)} := (\mathbb{E}_{x \in V} |f_{\text{str}}(x) - f_\varepsilon(x)|^2)^{1/2} \leq \varepsilon$.
- (f_{psd} pseudorandom) For every function $g : V \rightarrow {}^*\mathbb{C}$ that depends on only boundedly many functions from \mathcal{S} , one has $\langle f_{\text{psd}}, g \rangle_{L^2(V)} := \mathbb{E}_{x \in V} f_{\text{psd}}(x) \overline{g(x)} = o(1)$.

Proof. Given any finite subset \mathcal{S}_0 of \mathcal{S} , let $\mathcal{B}(\mathcal{S}_0)$ be the σ -algebra of V generated by the level sets of the functions of \mathcal{S}_0 ; this is a finite σ -algebra, with every atom being a limit subset of V . Given such a σ -algebra, we can define the conditional expectation $\mathbb{E}(f|\mathcal{B}(\mathcal{S}_0)) : V \rightarrow {}^*\mathbb{C}$ of f by the formula

$$\mathbb{E}(f|\mathcal{B}(\mathcal{S}_0))(x) := \mathbb{E}_{y \in \mathcal{B}(\mathcal{S}_0)(x)} f(y)$$

for all $x \in V$, where $\mathcal{B}(\mathcal{S}_0)(x)$ is the atom of $\mathcal{B}(\mathcal{S}_0)$ that contains x . Clearly $\mathbb{E}(f|\mathcal{B}(\mathcal{S}_0))$ is bounded in magnitude by A , and the *energy*

$$\mathcal{E}(\mathcal{S}_0) := \|\mathbb{E}(f|\mathcal{B}(\mathcal{S}_0))\|_{L^2(V)}^2$$

is a non-negative real number between 0 and A^2 . Let E_{max} denote the supremum of $\mathcal{E}(\mathcal{S}_0)$; as the energy is monotone in \mathcal{S}_0 we can thus (using the axiom of choice) find an increasing sequence \mathcal{S}_n for $n \in \mathbb{N}$ such that $\mathcal{E}(\mathcal{S}_n) \rightarrow E_{\text{max}}$.

From Pythagoras' theorem we have

$$\text{st}\|\mathbb{E}(f|\mathcal{B}(\mathcal{S}_{n'})) - \mathbb{E}(f|\mathcal{B}(\mathcal{S}_n))\|_{L^2(V)}^2 = \mathcal{E}(\mathcal{S}_{n'}) - \mathcal{E}(\mathcal{S}_n)$$

for any $n' > n$, where $\text{st}x$ denotes the standard part of the limit real x . Thus the $\mathbb{E}(f|\mathcal{B}(\mathcal{S}_n))$ are an L^2 Cauchy sequence in the sense that

$$\lim_{n, n' \rightarrow \infty} \text{st}\|\mathbb{E}(f|\mathcal{B}(\mathcal{S}_{n'})) - \mathbb{E}(f|\mathcal{B}(\mathcal{S}_n))\|_{L^2(V)}^2 = 0$$

We claim that this implies the existence of a limit function $f_{\text{str}} : V \rightarrow \mathbb{C}$, bounded in magnitude by A , such that

$$\lim_{n \rightarrow \infty} \text{st}\|f_{\text{str}} - \mathbb{E}(f|\mathcal{B}(\mathcal{S}_n))\|_{L^2(V)}^2 = 0.$$

Indeed, if we write $V = \prod_{\alpha \rightarrow \alpha_\infty} V_\alpha$, $f = \lim_{\alpha \rightarrow \alpha_\infty} f_\alpha$, and $\mathcal{S}_n = \prod_{\alpha \rightarrow \alpha_\infty} \mathcal{S}_{n, \alpha}$, one can set

$$f_{\text{str}} := \lim_{\alpha \rightarrow \alpha_\infty} \mathbb{E}(f_\alpha|\mathcal{B}(\mathcal{S}_{n_\alpha, \alpha}))$$

and the claim will follow if n_α increases to infinity at a sufficiently slow rate; we omit the routine details.

Now let g depend on a bounded number \mathcal{S}' of functions from \mathcal{S} , such that g is bounded in magnitude by B . Then for any standard natural n , one can rewrite

$$\text{st}\langle f - \mathbb{E}(f|\mathcal{B}(\mathcal{S}_n)), g \rangle_{L^2(V)}$$

as

$$\text{st}\langle \mathbb{E}(f|\mathcal{B}(\mathcal{S}_n \cup \mathcal{S}')) - \mathbb{E}(f|\mathcal{B}(\mathcal{S}_n)), g \rangle_{L^2(V)}$$

which by the Cauchy-Schwarz inequality is bounded in magnitude by

$$B\|\mathbb{E}(f|\mathcal{B}(\mathcal{S}_n \cup \mathcal{S}')) - \mathbb{E}(f|\mathcal{B}(\mathcal{S}_n))\|_{L^2(V)}$$

which by Pythagoras' theorem and definition of E_{\max} is bounded by

$$B(E_{\max} - \mathcal{E}(\mathcal{S}_n))^{1/2}$$

and thus

$$\lim_{n \rightarrow \infty} \text{st} \langle f - \mathbb{E}(f | \mathcal{B}(\mathcal{S}_n)), g \rangle_{L^2(V)} = 0.$$

Taking limits using the Cauchy-Schwarz and triangle inequalities, we conclude that

$$\lim_{n \rightarrow \infty} \text{st} \langle f - f_{\text{str}}, g \rangle_{L^2(V)} = 0.$$

Setting $f_{\text{psd}} := f - f_{\text{str}}$, we now obtain the claim. \square

Remark 11.2. Although we will not need this fact here, it is often useful to observe that if f is non-negative, then f_{str} and f_{ε} can be taken to be non-negative also. One can also establish this lemma using the machinery of *Loeb measure* [32]: if μ is Loeb measure on V , then f_{str} is essentially the conditional expectation (in $L^2(\mu)$) of f with respect to the σ -algebra generated by \mathcal{S} . See [37] for an implementation of this approach (and [44] for some further discussion of the role of Loeb measure in the nonstandard version of the Gowers norms).

Next, we recall one of the main theorems from [3], phrased in the ultralimit setting:

Theorem 11.3 (Weak inverse Gowers conjecture). *Let $s \geq 0$ be standard. Then there exists a standard integer $d = d(s, p) \geq 0$ such that for every limit finite-dimensional vector space V and every bounded limit function $f : V \rightarrow {}^*\mathbb{C}$ with $\|f\|_{U^{s+1}(V)} \gg 1$, there exists $P \in \text{Poly}_{\leq d}(V \rightarrow {}^*\mathbb{T})$ such that $|\mathbb{E}_{x \in V} f(x) e(-P(x))| \gg 1$.*

Proof. See [3, Corollary 1.23]. The translation to the ultralimit setting proceeds exactly as in Section 3. \square

We are now ready to deduce Theorem 1.11 from Theorem 1.20. Fix s, p , and let d be the minimal integer for which Theorem 11.3 holds for this value of s, p . If $d \leq s$, then we are done by Theorem 3.2, so suppose for sake of contradiction that $d > s$.

By construction of d , we can find a limit finite-dimensional vector space V and a bounded limit function $f : V \rightarrow {}^*\mathbb{C}$ such that $\|f\|_{U^{s+1}(V)} \gg 1$, but such that

$$\langle f, e(P) \rangle_{L^2(V)} = o(1)$$

whenever P is a polynomial of degree $\leq d-1$. By Fourier analysis (and Lemma 1.7(vi)), this implies that

$$\langle f, g \rangle_{L^2(V)} = o(1)$$

whenever g is a function of a bounded number of polynomials of degree $\leq d-1$. In particular we have

$$\langle f, e(P) \rangle_{L^2(V)} = o(1)$$

whenever P is a polynomial of degree $\leq d$ of bounded rank.

Using Lemma 11.1 (and Lemma 1.7(vi)), we can decompose $f = f_{\text{str}} + f_{\text{psd}}$, where $f_{\text{str}}, f_{\text{psd}} : V \rightarrow {}^*\mathbb{C}$ are bounded limit functions, we have

$$\langle f_{\text{psd}}, g \rangle_{L^2(V)} = o(1)$$

whenever g is a function of a bounded number of polynomials of degree $\leq d$, and for every $\varepsilon > 0$ we can approximate f_{str} by a bounded limit function f_ε that is a function of a bounded number of polynomials of degree $\leq d$. In particular, by Theorem 11.3, one has

$$\|f_{\text{psd}}\|_{U^{s+1}(V)} = o(1)$$

and hence by the triangle inequality for $U^{s+1}(V)$ (Lemma B.1(i)) one has

$$\|f_{\text{str}}\|_{U^{s+1}(V)} \gg 1.$$

Also from the triangle inequality we see that

$$\langle f_{\text{str}}, e(P) \rangle_{L^2(V)} = o(1) \tag{11.1}$$

whenever P is a polynomial of degree $\leq d$ of bounded rank.

As the $f_{\text{str}} - f_\varepsilon$ are uniformly bounded in ε , and have (the standard part of the) L^2 norm going to zero as $\varepsilon \rightarrow 0$, we see (using Lemma B.1(iii)) that

$$\text{st}\|f_{\text{str}} - f_\varepsilon\|_{U^{s+1}(V)} \rightarrow 0$$

as $\varepsilon \rightarrow 0$. For all sufficiently small $\varepsilon > 0$, we thus have

$$\|f_\varepsilon\|_{U^{s+1}(V)} \gg 1 \tag{11.2}$$

uniformly in ε .

By Fourier analysis, we can express f_ε as a bounded linear combination of phases $e(P)$, where the P are polynomials of degree $\leq d$. We separate $f_\varepsilon = f'_\varepsilon + f''_\varepsilon$, where f'_ε is a linear combination of phases $e(P)$ of unbounded rank, and f''_ε is a linear combination of phases of $e(P)$ bounded rank.

From Theorem 1.20 (and Theorem 3.1) we see that $\|e(P)\|_{U^d(V)} = o(1)$ whenever P has unbounded rank, and in particular (by Lemma B.1(ii)) $\|e(P)\|_{U^{s+1}(V)} = o(1)$ and $\mathbb{E}_{x \in V} e(P(x)) = o(1)$. Since the difference of a degree d polynomial of unbounded rank and a degree d polynomial of bounded rank remains of unbounded rank, we also have $\mathbb{E}_{x \in V} e(P(x) - Q(x)) = o(1)$ whenever Q is of bounded rank. We conclude that $f'_\varepsilon, f''_\varepsilon$ are essentially orthogonal in the sense that

$$\langle f'_\varepsilon, f''_\varepsilon \rangle_{L^2(V)} = o(1)$$

and hence

$$\|f''_\varepsilon\|_{L^2(V)}^2 = \langle f_\varepsilon, f''_\varepsilon \rangle_{L^2(V)} + o(1).$$

On the other hand, from (11.1) one has

$$\langle f_{\text{str}}, f''_\varepsilon \rangle_{L^2(V)} = o(1)$$

while from the Cauchy-Schwarz inequality we see that

$$\text{st}\langle f_{\text{str}} - f_\varepsilon, f''_\varepsilon \rangle_{L^2(V)} \rightarrow 0$$

as $\varepsilon \rightarrow 0$. We conclude that

$$\text{st}\|f''_\varepsilon\|_{L^2(V)} \rightarrow 0$$

as $\varepsilon \rightarrow 0$, which in particular implies (by Lemma B.1(iii)) that

$$\text{st}\|f''_\varepsilon\|_{U^{s+1}(V)} \rightarrow 0$$

as $\varepsilon \rightarrow 0$. Also, as f'_ε is a bounded linear combination of $e(P)$ for P of unbounded rank, and thus of infinitesimal $U^{s+1}(V)$ norm, we see from the triangle inequality (Lemma B.1(i)) that

$$\|f'_\varepsilon\|_{U^{s+1}(V)} = o(1)$$

and hence

$$\text{st}\|f_\varepsilon\|_{U^{s+1}(V)} \rightarrow 0$$

as $\varepsilon \rightarrow 0$, contradicting (11.2). This concludes the deduction of Theorem 1.11 from Theorem 1.20.

APPENDIX A. BASIC THEORY OF ULTRALIMITS

In this appendix we review the machinery of ultralimits.

We will assume the existence of a *standard universe* \mathfrak{U} which contains all the objects and spaces of interest for Theorem 1.11 or Theorem 1.20, such as the natural numbers, standard finite-dimensional vector spaces \mathbb{F}^n and their elements, the unit circle \mathbb{T} and its elements, functions from the former spaces to the latter (such as polynomials $P \in \text{Poly}_{\leq d}(\mathbb{F}^n \rightarrow \mathbb{T})$), and so forth. The precise construction of this universe is not important, so long as it forms a set. We refer to objects and spaces inside the standard universe as *standard objects* and *standard spaces*, with the latter being sets whose elements are in the former category. Thus for instance, elements of \mathbb{N} are standard natural numbers, and for every standard natural number n , \mathbb{F}^n is a standard finite-dimensional vector space. Strictly speaking, the universe \mathfrak{U} cannot contain *all* finite-dimensional vector spaces, as the class of such spaces is not a set, but for the purposes of proving Theorem 1.11 or Theorem 1.20 we only need to pick one representative of each isomorphism class of such spaces, such as \mathbb{F}^n for $n = 0, 1, 2, \dots$, and these certainly form a set.

The one technical ingredient we need is the following:

Lemma A.1 (Ultrafilter lemma). *There exists a collection α of subsets of the natural numbers \mathbb{N} with the following properties:*

- (i) (*Monotonicity*) If $A \in \alpha$ and $B \supset A$, then $B \in \alpha$.
- (ii) (*Closure under intersection*) If $A, B \in \alpha$, then $A \cap B \in \alpha$.
- (iii) (*Maximality*) If $A \subset \mathbb{N}$, then either $A \in \alpha$ or $\mathbb{N} \setminus A \in \alpha$, but not both.
- (iv) (*Non-principality*) If $A \in \alpha$, and A' is formed from A by adding or deleting finitely many elements to or from A , then $A' \in \alpha$.

Proof. The collection of subsets of \mathbb{N} which are cofinite (i.e. whose complement is finite) already obeys the monotonicity, closure under intersection, and non-principality

properties. Using Zorn's lemma³, one can enlarge this collection to a maximal collection, which then obeys all the required properties. \square

Throughout the paper, we fix a non-principal ultrafilter α . A property $P(\mathbf{n})$ depending on a natural number \mathbf{n} is said to hold *for \mathbf{n} sufficiently close to α* if the set of \mathbf{n} for which $P(\mathbf{n})$ holds lies in α .

Once we have fixed this ultrafilter, we can now define limit objects and spaces:

Definition A.2 (Limit objects). Given a sequence $(x_{\mathbf{n}})_{\mathbf{n} \in \mathbb{N}}$ of standard objects in \mathfrak{U} , we define their *ultralimit* $\lim_{\mathbf{n} \rightarrow \alpha} x_{\mathbf{n}}$ to be the equivalence class of all sequences $(y_{\mathbf{n}})_{\mathbf{n} \in \mathbb{N}}$ of standard objects in \mathfrak{U} such that $x_{\mathbf{n}} = y_{\mathbf{n}}$ for \mathbf{n} sufficiently close to α . Note that the ultralimit $\lim_{\mathbf{n} \rightarrow \alpha} x_{\mathbf{n}}$ can also be defined even if $x_{\mathbf{n}}$ is only defined for \mathbf{n} sufficiently close to α .

An ultralimit of standard natural numbers is known as a *limit natural number*, an ultralimit of standard real numbers is known as a *limit real number*, etc.

For any standard object x , we identify x with its own ultralimit $\lim_{\mathbf{n} \rightarrow \alpha} x$. Thus, every standard natural number is a limit natural number, etc.

Any operation or relation on standard objects can be extended to limit objects in the obvious manner. For instance, the sum of two limit real numbers $\lim_{\mathbf{n} \rightarrow \alpha} x_{\mathbf{n}}$, $\lim_{\mathbf{n} \rightarrow \alpha} y_{\mathbf{n}}$ is the limit real number

$$\lim_{\mathbf{n} \rightarrow \alpha} x_{\mathbf{n}} + \lim_{\mathbf{n} \rightarrow \alpha} y_{\mathbf{n}} = \lim_{\mathbf{n} \rightarrow \alpha} (x_{\mathbf{n}} + y_{\mathbf{n}}),$$

and the statement $\lim_{\mathbf{n} \rightarrow \alpha} x_{\mathbf{n}} < \lim_{\mathbf{n} \rightarrow \alpha} y_{\mathbf{n}}$ means that $x_{\mathbf{n}} < y_{\mathbf{n}}$ for all \mathbf{n} sufficiently close to α .

Remark A.3. A famous theorem of Łoś asserts that any statement in first-order logic which is true about standard objects, is automatically true for limit objects as well. For instance, the standard real numbers form an ordered field, and so the limit real numbers do also, because the axioms of an ordered field can be phrased in first-order logic. We will use this theorem in the sequel without further comment.

Definition A.4 (Limit spaces and functions). Let $(X_{\mathbf{n}})_{\mathbf{n} \in \mathbb{N}}$ be a sequence of standard spaces $X_{\mathbf{n}}$ in \mathfrak{U} indexed by the natural numbers. The *ultraproduct* $\prod_{\mathbf{n} \rightarrow \alpha} X_{\mathbf{n}}$ of the $X_{\mathbf{n}}$ is defined to be the space of all ultralimits $\lim_{\mathbf{n} \rightarrow \alpha} x_{\mathbf{n}}$, where $x_{\mathbf{n}} \in X_{\mathbf{n}}$ for all \mathbf{n} . Note $X_{\mathbf{n}}$ only needs to be well-defined for \mathbf{n} sufficiently close to α in order for the ultraproduct to be well-defined. If X is a set, the set $\prod_{\mathbf{n} \rightarrow \alpha} X$ is known as the *ultrapower* of X and

³By using this lemma, our results thus rely on the axiom of choice, which we will of course assume throughout this paper. On the other hand, it is possible to rephrase Theorem 1.11 and Theorem 1.20 in the language of Peano arithmetic. Applying a famous theorem of Gödel[10], we then conclude that Theorem 1.20 is provable in ZFC if and only if it is provable in ZF. In fact, it is possible (with some effort) to directly translate these ultrafilter arguments to a (lengthier) argument in which ultrafilters or the axiom of choice is not used. We will not do so here, though, as the translation is quite tedious. In particular, the regularity lemma and equidistribution arguments in this paper will become messier, resembling those that appear in [22] or [28].

is denoted *X . Thus for instance ${}^*\mathbb{N}$ is the space of all limit natural numbers, ${}^*\mathbb{R}$ is the space of all limit reals, etc.

We define a *limit set* to be an ultraproduct of sets, a *limit group* to be an ultraproduct of groups, a *limit finite set* to be an ultraproduct of finite sets, and so forth. A *limit subset* of a limit set $X = \prod_{\mathbf{n} \rightarrow \alpha} X_{\mathbf{n}}$ is a limit set of the form $Y = \prod_{\mathbf{n} \rightarrow \alpha} Y_{\mathbf{n}}$, where $Y_{\mathbf{n}}$ is a standard subset of $X_{\mathbf{n}}$ for all \mathbf{n} sufficiently close to α .

Given a sequence of standard functions $f_{\mathbf{n}} : X_{\mathbf{n}} \rightarrow Y_{\mathbf{n}}$ between standard sets $X_{\mathbf{n}}, Y_{\mathbf{n}}$, we can form the *ultralimit* $f = \lim_{\mathbf{n} \rightarrow \alpha} f_{\mathbf{n}}$ to be the function $f : \prod_{\mathbf{n} \rightarrow \alpha} X_{\mathbf{n}} \rightarrow \prod_{\mathbf{n} \rightarrow \alpha} Y_{\mathbf{n}}$ defined by the formula

$$f(\lim_{\mathbf{n} \rightarrow \alpha} x_{\mathbf{n}}) := \lim_{\mathbf{n} \rightarrow \alpha} f_{\mathbf{n}}(x_{\mathbf{n}}).$$

We refer to f as a *limit function* or *limit map*.

Remark A.5. In the nonstandard analysis literature, limit natural numbers are known as *nonstandard natural numbers*, limit sets are known as *internal sets*, and limit functions are known as *internal functions*. We have chosen the limit terminology instead as we believe that it is less confusing and emphasises the role of ultralimits in the subject.

It is important to note that not every subset of a limit set is again a limit set, for instance \mathbb{N} is not a limit subset of ${}^*\mathbb{N}$ (this fact is known as the *overspill principle*). Indeed, one can think of the limit subsets of a limit set as being analogous to the measurable subsets of a measure space. In a similar vein, not every function between two limit sets is a limit function; in this regard, limit functions are analogous to measurable functions. This analogy can be deepened by using the theory of *Loeb measures*, but we will avoid using this machinery here.

A.6. Asymptotic notation. By taking ultralimits, one can formalise asymptotic notation, such as the $O()$ notation, in a manner that requires no additional quantifiers:

Definition A.7 (Asymptotic notation). A limit complex number X is said to be *bounded* if one has $|X| \leq C$ for some standard real number C , in which case we also write $X = O(1)$ or $|X| \ll 1$. More generally, given a limit complex number X and limit non-negative number Y , we write $|X| \ll Y$, $Y \gg |X|$, or $X = O(Y)$ if one has $|X| \leq CY$ for some standard real number C . We write $X = o(Y)$ if one has $|X| \leq \varepsilon Y$ for every standard $\varepsilon > 0$. Observe that for any X, Y with Y positive, one has either $|X| \gg Y$ or $X = o(Y)$. We say that X is *infinitesimal* if $X = o(1)$, and *unbounded* if $1/X = o(1)$. Thus for instance any limit complex number X will either be bounded or unbounded.

Example A.8. The limit real $\lim_{\mathbf{n} \rightarrow \alpha} 1/\mathbf{n}$ defines an infinitesimal, but non-zero, limit real number x ; its reciprocal $\lim_{\mathbf{n} \rightarrow \alpha} \mathbf{n}$ is an unbounded limit real.

From the Bolzano-Weierstrass theorem, every bounded limit complex number x can be expressed uniquely as the sum of a standard real number $\text{st}(x)$ and an infinitesimal $x - \text{st}(x)$; we refer to $\text{st}(x)$ as the *standard part* of x .

APPENDIX B. PROPERTIES OF THE GOWERS NORMS

In this appendix we record some basic properties of the Gowers norms. We use the normalised L^p norms

$$\|f\|_{L^p(G)} := (\mathbb{E}_{x \in G} |f(x)|^p)^{1/p}$$

for any finite non-empty set G and any $f : G \rightarrow \mathbb{C}$.

Lemma B.1. *Let $G = (G, +)$ be a finite abelian group, and let $d \geq 1$ be an integer.*

- (i) *The Gowers norm $\|\cdot\|_{U^d(G)}$ is a norm on functions $f : G \rightarrow \mathbb{C}$ for $d \geq 2$, and a semi-norm for $d = 1$. In particular, we have the Gowers triangle inequality*

$$\|f + g\|_{U^d(G)} \leq \|f\|_{U^d(G)} + \|g\|_{U^d(G)}$$

for $f, g : G \rightarrow \mathbb{C}$.

- (ii) *One has the monotonicity property*

$$\|f\|_{U^d(G)} \leq \|f\|_{U^{d+1}(G)}$$

for all $f : G \rightarrow \mathbb{C}$. In particular

$$|\mathbb{E}_{x \in V} f(x)| = \|f\|_{U^1(G)} \leq \|f\|_{U^d(G)}.$$

- (iii) *One has the bound*

$$\|f\|_{U^d(G)} \leq \|f\|_{L^{2^d/(d+1)}(G)}$$

for all $f : G \rightarrow \mathbb{C}$.

- (iv) *One has the first Cauchy-Schwarz-Gowers inequality*

$$|\mathbb{E}_{x, h_1, \dots, h_d \in G} \prod_{\omega \in \{0,1\}^d} f_\omega(x + \omega_1 h_1 + \dots + \omega_d h_d)| \leq \prod_{\omega \in \{0,1\}^d} \|f_\omega\|_{U^d(G)}$$

for all $\{0,1\}^d$ -tuples $(f_\omega)_{\omega \in \{0,1\}^d}$ of functions $f_\omega : G \rightarrow \mathbb{C}$, where $\omega := (\omega_1, \dots, \omega_d)$.

- (v) *One has the second Cauchy-Schwarz-Gowers inequality*

$$|\mathbb{E}_{x_1, \dots, x_d \in G} f(x_1 + \dots + x_d) \prod_{j=1}^d F_j(x_1, \dots, x_d)| \leq \|f\|_{U^d(G)}$$

for all $f : G \rightarrow \mathbb{C}$ and $F_j : G^d \rightarrow \mathbb{C}$, if each F_j is bounded in magnitude by 1 and is independent of the x_j variable.

- (vi) *If $P \in \text{Poly}_{\leq d-1}(G \rightarrow \mathbb{T})$ and $f : G \rightarrow \mathbb{C}$ then*

$$\|fe(P)\|_{U^d(G)} = \|f\|_{U^d(G)}. \tag{B.1}$$

Proof. Claim (i) is proven in [12, Lemma 3.9], [17, Section 5.1] or [42, Section 11.1]. Claim (ii) is proven in [42, Section 11.1]. Claim (iii) (which is also [42, Exercise 11.1.13]) follows easily from the recursive formula

$$\|f\|_{U^{d+1}(V)} = (\mathbb{E}_{h \in V} \|\Delta_h f\|_{U^d(V)}^{2^d})^{1/2^{d+1}}$$

and induction on d , together with the special case

$$(\mathbb{E}_{h \in V} \|f(\cdot + h)g(\cdot)\|_{L^{2^d/(d+1)}(V)}^{2^d})^{1/2^d} \leq \|f\|_{L^{2^{d+1}/(d+2)}(V)} \|g\|_{L^{2^{d+1}/(d+2)}(V)}$$

of Young's convolution inequality, which follows from the more traditional instance

$$(\mathbb{E}_{h \in V} |\mathbb{E}_{x \in V} F(x+h)G(h)|^{d+1})^{1/(d+1)} \leq \|F\|_{L^{2(d+1)/(d+2)}(V)}$$

of that inequality by setting $F := |f|^{2^d/(d+1)}$ and $G := |g|^{2^d/(d+1)}$.

Claim (iv) is proven in [12, Lemma 3.8], [17, Section 5.1] or [42, Section 11.1]; Claim (v) is proven in [19, Appendix B]. Claim (vi) follows immediately from the identity $\Delta_{h_1} \dots \Delta_{h_d} e(P) = 1$ for all $h_1, \dots, h_d \in V$. \square

APPENDIX C. POLYNOMIAL ALGEBRA

In this appendix we review the general theory of polynomial maps (and related objects, such as cubes) on (filtered) groups that are not necessarily abelian (in particular, they may be nilpotent). This theory was initiated by Lazard [29] and Leibman [30], [31] (inspired in part by the classical Hall-Petresco formula [25], [34]), and further developed by Host and Kra [26], [27] and by Green and the authors [19], [24]. Our discussion here is largely drawn from the paper [24].

Polynomial algebra works on both multiplicative groups $G = (G, \cdot)$ and on additive groups $G = (G, +)$. For sake of concreteness we shall set out the theory here using multiplicative group notation, but one can of course adapt all the definitions here to additive groups in an obvious manner, and in fact most of the applications of this theory in this paper will be in the additive setting. Our conventions will be that additive groups are always understood to be abelian, whereas multiplicative groups are not necessarily abelian.

The concepts here can be defined both in the standard and nonstandard setting, but again for concreteness we shall work purely in the standard universe in this appendix. But all of the results here can easily be phrased in the language of first-order logic (they involve only finitely many quantifiers) and so extend without difficulty to the nonstandard universe also.

C.1. The category of filtered groups. The theory of polynomial maps is most naturally expressed in terms of a certain category of filtered groups, which we will now define.

Definition C.2 (Filtered group). A *filtered group* $G = (G, G_{\mathbb{N}})$ is a multiplicative group $G = (G, \cdot)$, together with a nested sequence

$$G \geq G_0 \geq G_1 \geq \dots$$

of subgroups $G_{\mathbb{N}} = (G_n)_{n \in \mathbb{N}}$, obeying the commutator relation $[G_i, G_j] \subset G_{i+j}$ for all $i, j \in \mathbb{N}$, where $[G_i, G_j]$ is the group generated by the commutators $g_i^{-1} g_j^{-1} g_i g_j$ with $g_i \in G_i, g_j \in G_j$. We refer to $G_{\mathbb{N}}$ as a *filtration* of G .

A filtered group is said to have *degree* $\leq s$ for some natural number s if G_i is trivial for all $i > s$.

Example C.3 (Abelian case). When the group $G = (G, +)$ is additive (and thus abelian), a filtered group is simply a nested sequence $G \geq G_0 \geq G_1 \geq \dots$ of subspaces (since the commutator relation is automatic in this case). In particular, for any natural number $k \geq 0$, one can give any additive group G the *maximal degree k filtration* $G_{\mathbb{N}} = G_{\mathbb{N}}^{(k)}$, defined by setting G_i equal to G when $i \leq k$ and $G_i = \{0\}$ for $i > k$.

Example C.4 (Lower central series). Any group G can become a filtered group by taking G_i to be the lower central series of G , thus $G_0 = G_1 := G$ and $G_{i+1} := [G, G_i]$ for $i \geq 1$.

Remark C.5. In this paper we will only filter groups G by the natural numbers \mathbb{N} . However it is sometimes convenient to filter groups by other sets, such as \mathbb{N}^k , in order to develop a theory of “multidegree” for polynomials of several variables; see [24]. While one could use this notation to describe the multilinear maps that arise in this paper, we have chosen not to do so here in order not to add even more terminology to what is already quite a notation-intensive argument.

An obvious way to make the class of all filtered groups a category is to use the *filtered homomorphisms* $\phi : H \rightarrow G$ between two filtered groups $H = (H, H_{\mathbb{N}})$, $G = (G, G_{\mathbb{N}})$, defined as a group homomorphism from H to G that maps each H_i to G_i . However, this turns out to be too small a class of morphisms for our purposes, and we will need to use instead the larger class of *polynomial maps* between two filtered groups. This concept can be defined in a number of different ways. The quickest way is via differentiation:

Definition C.6 (Polynomial maps via differentiation). Let $H = (H, H_{\mathbb{N}})$, $G = (G, G_{\mathbb{N}})$ be filtered groups, and let $\phi : H \rightarrow G$ be a map. For any $h \in H$, we define the *derivative* $\partial_h \phi : H \rightarrow G$ of ϕ in the direction h by the formula

$$\partial_h \phi(x) := \phi(hx)\phi(x)^{-1}.$$

We say that the map $\phi : H \rightarrow G$ is a *polynomial map* if one has

$$\partial_{h_1} \dots \partial_{h_m} \phi(x) \in G_{i_1 + \dots + i_m}$$

whenever $m \geq 0$ and $i_1, \dots, i_m \in \mathbb{N}$, and $h_j \in H_{i_j}$ for all $1 \leq j \leq m$. The space of all polynomial maps from H to G will be denoted $\text{Poly}(H \rightarrow G)$.

Example C.7 (Non-classical polynomials as polynomial maps). If V, G are additive groups, with V given the maximal degree ≤ 1 filtration, and G the maximal degree $\leq k$ filtration for some $k \in \mathbb{N}$, then $\text{Poly}(V \rightarrow G)$ corresponds precisely to the space $\text{Poly}_{\leq k}(V \rightarrow G)$ defined in Definition 1.2 (this definition was for finite-dimensional vector spaces V , but the definition clearly also makes sense for other additive groups). In particular, a non-classical polynomial $P : V \rightarrow \mathbb{T}$ of degree $\leq k$ is also a polynomial map from V (with the maximal degree ≤ 1 filtration) to \mathbb{T} (with the maximal degree $\leq k$ filtration).

Example C.8. Every filtered homomorphism is a polynomial map. For any $g \in G$, the left translation maps $x \mapsto gx$ and right translation maps $x \mapsto xg$ are polynomial maps from G to itself.

Remark C.9. A basic theorem of Lazard and Leibman [29], [30], [31] asserts that $\text{Poly}(H \rightarrow G)$ is a group; see e.g. [24, Corollary B.11]. This generalises the (obvious) fact that $\text{Poly}_{\leq k}(V \rightarrow G)$ is a group in the additive case.

A convenient fact about polynomiality is that it suffices to check it on generators:

Proposition C.10 (Checking polynomiality on generators). *Let H, G be filtered groups, and for each $i \in \mathbb{N}$, let E_i be a set of generators for H_i . Then a map $\phi : H \rightarrow G$ is polynomial if and only if*

$$\partial_{h_1} \dots \partial_{h_m} \phi(x) \in G_{i_1 + \dots + i_m}$$

whenever $m \geq 0$ and $i_1, \dots, i_m \in \mathbb{N}$, and $h_j \in E_{i_j}$ for all $1 \leq j \leq m$.

Proof. See [24, Proposition B.17]. □

It is not immediately obvious from Definition C.6 that the polynomial maps turn the class of filtered groups into a category, because one has to show that the composition of two polynomial maps is still polynomial. However, this can be achieved via the machinery of *Host-Kra cube groups*, which we now pause to define.

Definition C.11 (Host-Kra cube group). Let $G = (G, G_{\mathbb{N}})$ be a filtered group. For any $k \in \mathbb{N}$, we define the k^{th} *Host-Kra cube group* $\text{HK}^k(G) = \text{HK}^k(G, G_{\mathbb{N}})$ of this filtration to be the subgroup of $G^{\{0,1\}^k}$ generated by those elements $(g_{\omega})_{\omega \in \{0,1\}^k}$ which take the form $g_{\omega} = g$ for $\omega \in F$ and $g_{\omega} = \text{id}$ otherwise, where F is a face of $\{0,1\}^k$ of some codimension i , and g is an element of G_i . Elements $(g_{\omega})_{\omega \in \{0,1\}^k}$ of $\text{HK}^k(G)$ will be referred to as *k -dimensional cubes*⁴ in G .

We have an alternate description of these groups via a “Taylor expansion”:

Proposition C.12. *Let $G = (G, G_{\mathbb{N}})$ be a filtered group, let $k \in \mathbb{N}$, and let $g \in G^{\{0,1\}^k}$. Then $g \in \text{HK}^k(G)$ if and only if there exist “Taylor coefficients” $g_J \in G_{|J|}$ for each subset $J \subset \{1, \dots, k\}$ such that*

$$g = \left(\prod_{J \subset \{1, \dots, k\}} g_J^{\prod_{j \in J} \omega_j} \right)_{\omega \in \{0,1\}^k} = \prod_{J \subset \{1, \dots, k\}} (g_J^{\prod_{j \in J} \omega_j})_{\omega \in \{0,1\}^k}, \quad (\text{C.1})$$

where the subsets of $\{1, \dots, k\}$ are ordered lexicographically (i.e. $J < J'$ whenever $\sum_{j \in J} 2^{-j} < \sum_{j \in J'} 2^{-j}$). Furthermore, the g_J are determined uniquely by g .

Proof. See [21, Lemma 6.4]. □

Thus, for instance, $\text{HK}^2(G)$ consists of all tuples of the form $(g_{00}, g_{00}g_{01}, g_{00}g_{10}, g_{00}g_{01}g_{10}g_{11})$, where $g_{00} \in G_0$, $g_{01}, g_{10} \in G_1$, and $g_{11} \in G_2$.

Theorem C.13 (Polynomial maps via cubes). *Let $G = (G, G_{\mathbb{N}})$ and $H = (H, H_{\mathbb{N}})$ be filtered groups, and let $\phi : H \rightarrow G$ be a map. Then ϕ is a polynomial map if and only if ϕ preserves cubes, in the sense that for any $k \in \mathbb{N}$ and $(h_{\omega})_{\omega \in \{0,1\}^k} \in \text{HK}^k(H)$, the tuple $(\phi(h_{\omega}))_{\omega \in \{0,1\}^k}$ lies in $\text{HK}^k(G)$. (In other words, the map $\phi^{\oplus \{0,1\}^k} : H^{\{0,1\}^k} \rightarrow G^{\{0,1\}^k}$ maps $\text{HK}^k(H)$ to $\text{HK}^k(G)$.)*

⁴These cubes are also referred to as *parallelopeds* in some literature, e.g. [27].

As an immediate corollary of this theorem, we see that the composition of two polynomial maps is again polynomial, and so the class of filtered maps is now a category. If $\phi : H \rightarrow G$ is a polynomial map and $k \in \mathbb{N}$, we use $\text{HK}^k(\phi) : \text{HK}^k(H) \rightarrow \text{HK}^k(G)$ to denote the restriction of $\phi^{\oplus\{0,1\}^k} : H^{\{0,1\}^k} \rightarrow G^{\{0,1\}^k}$ to $\text{HK}^k(H)$ and $\text{HK}^k(G)$.

Proof. When H is additive, this theorem was proven in [21, Proposition 6.5]. We will give an alternate proof based on [24, Theorem B.10]. To use this theorem, we need a generalisation of the Host-Kra groups. For any natural numbers $i_1, \dots, i_k \in \mathbb{N}$, define the *Host-Kra group* $\text{HK}^{i_1, \dots, i_k}(G)$ of a filtered group G to be the subgroup of $G^{\{0,1\}^k}$ generated by the elements of the form

$$(g_\omega)_{\omega \in \{0,1\}^k},$$

where $\omega_0 \in \{0,1\}^k$, $g_{\omega_0} \in G_{\sum_{(\omega_0)_j=1} i_j}$, and g_ω equals g_{ω_0} when $\omega_j \geq (\omega_0)_j$ for all $1 \leq j \leq k$, and is the identity otherwise. Thus for instance, when $i_1 = \dots = i_k = 1$, then $\text{HK}^{1, \dots, 1}(G) = \text{HK}^k(G)$. It is easy to adapt the proof of Proposition C.12 to see that elements of $\text{HK}^{i_1, \dots, i_k}(G)$ are precisely those tuples of the form (C.1), where each Taylor coefficient g_J now lies in $G_{\sum_{j \in J} i_j}$ rather than $G_{|J|}$.

The result [24, Theorem B.10] asserts that ϕ is a polynomial map if and only if ϕ (or more precisely, $\phi^{\oplus\{0,1\}^k}$) maps $\text{HK}^{i_1, \dots, i_k}(H)$ to $\text{HK}^{i_1, \dots, i_k}(G)$ for every $k, i_1, \dots, i_k \in \mathbb{N}$. In view of this result, to prove Theorem C.13, it suffices to show that if ϕ maps $\text{HK}^k(H)$ to $\text{HK}^k(G)$, then it maps $\text{HK}^{i_1, \dots, i_k}(H)$ to $\text{HK}^{i_1, \dots, i_k}(G)$ for every $k, i_1, \dots, i_k \in \mathbb{N}$.

For any $k, i_1, \dots, i_k \in \mathbb{N}$. Let $P(i_1, \dots, i_k)$ denote the assertion that $\phi^{\oplus\{0,1\}^k}$ maps $\text{HK}^{i_1, \dots, i_k}(H)$ to $\text{HK}^{i_1, \dots, i_k}(G)$. By hypothesis, $P(1, \dots, 1)$ is true for any number of 1's; our task is to then show that $P(i_1, \dots, i_k)$ is true in general. The case $k = 0$ is trivial, so we may assume that $k \geq 1$.

Suppose first that $i_k = 0$. An inspection of the definition then shows that $\text{HK}^{i_1, \dots, i_{k-1}, 0}(G) = \text{HK}^{i_1, \dots, i_{k-1}}(G) \times \text{HK}^{i_1, \dots, i_{k-1}}(G)$, and similarly for H . As a consequence, we see that $P(i_1, \dots, i_{k-1}, 0)$ is implied by $P(i_1, \dots, i_{k-1})$. From this observation we may assume without loss of generality that $i_k \geq 1$. By symmetry, we may in fact assume that $i_j \geq 1$ for $j = 1, \dots, k$.

Consider the map $\eta : G^{\{0,1\}^k} \rightarrow G^{\{0,1\}^{k+i_k-1}}$ defined by

$$\eta((g_\omega)_{\omega \in \{0,1\}^k}) := (g_{\omega_1, \dots, \omega_{k-1}, \min(\omega_k, \dots, \omega_{k+i_k-1})})_{\omega \in \{0,1\}^{k+i_k-1}}. \quad (\text{C.2})$$

This is clearly an injective group homomorphism when $i_k \geq 1$. We claim that

$$\eta(\text{HK}^{i_1, \dots, i_k}(G)) = \eta(G^{\{0,1\}^k}) \cap \text{HK}^{\tilde{i}_1, \dots, \tilde{i}_{k+i_k-1}}(G), \quad (\text{C.3})$$

where \tilde{i}_j is equal to i_j for $j < k$ and equal to 1 for $k \leq j \leq k + i_k - 1$. It is easy to see that the left-hand group in (C.3) is included in the right-hand side, simply by checking what η does to each generator of $\text{HK}^{i_1, \dots, i_k}(G)$. The reverse inclusion is a little trickier. Suppose that $g \in G^{\{0,1\}^k}$ is such that $\eta(g)$ lies in $\text{HK}^{\tilde{i}_1, \dots, \tilde{i}_{k+i_k-1}}(G)$. From (C.1) and induction, we see that the Taylor coefficients $\eta(g)_J$ of $\eta(g)$ vanish unless J either contains $\{k, \dots, k + i_k - 1\}$ or is disjoint from $\{k, \dots, k + i_k - 1\}$. As a consequence,

each of the factors $(\eta(g)_J^{\prod_{j \in J} \omega_j})_{\omega \in \{0,1\}^{k+i_k-1}}$ of the Taylor expansion of $\eta(g)$ are equal to $\eta(g_F)$ for some generator g_F of $\mathrm{HK}^{i_1, \dots, i_k}(G)$, and the claim follows.

From (C.3), we see that $P(i_1, \dots, i_k)$ is implied by $P(\tilde{i}_1, \dots, \tilde{i}_{k+i_k}) = P(i_1, \dots, i_{k-1}, 1, \dots, 1)$, where i_k copies of 1 appear in the latter expression. By symmetry and iteration, we conclude that $P(i_1, \dots, i_k)$ is implied by $P(1, \dots, 1)$ where $i_1 + \dots + i_k$ copies of 1 appear in the latter expression, and Theorem C.13 follows. \square

Remark C.14. The group $\mathrm{HK}^k(G)$ itself comes with a natural filtration, with $\mathrm{HK}^k(G)_i$ defined to be the Host-Kra group $\mathrm{HK}^k(G, (G_{j+i})_{j \in \mathbb{N}})$ of G with the shifted filtration $(G_{j+i})_{j \in \mathbb{N}}$; see [24, Proposition B.15]. Theorem C.13 can then be used to show that HK^k can be viewed as a functor from the category of filtered groups to itself. These functors are related to each other by the pleasant identity $\mathrm{HK}^j \circ \mathrm{HK}^k = \mathrm{HK}^{j+k}$ for all $j, k \in \mathbb{N}$; in particular, one can define HK^k recursively as an iteration of the functor HK^1 . We will however not adopt this perspective here.

C.15. The additive case. Let $V = (V, \mathbb{N})$ be an additive filtered group, and let $k \in \mathbb{N}$. By Proposition C.12, $\mathrm{HK}^k(V)$ consists precisely of those tuples of the form

$$\left(\sum_{J \subset \{1, \dots, k\}} \left(\prod_{j \in J} \omega_j \right) v_J \right)_{\omega \in \{0,1\}^k}$$

with $v_J \in V_{|J|}$ for all $J \subset \{1, \dots, k\}$. Thus, for instance, $\mathrm{HK}^2(V)$ is the space of all quadruples

$$(v_{00}, v_{00} + v_{01}, v_{00} + v_{10}, v_{00} + v_{01} + v_{10} + v_{11})$$

where $v_{00} \in V_0$, $v_{01}, v_{10} \in V_1$, and $v_{11} \in V_2$.

There is also an equivalent “dual” description of this space (which we will need to prove Proposition 9.8):

Proposition C.16 (Description of $\mathrm{HK}^k(V)$). *Let $V = (V, V_{\mathbb{N}})$ be an additive filtered group. Then for any $k \in \mathbb{N}$, $\mathrm{HK}^k(V)$ consists precisely of those tuples $(v_{\omega})_{\omega \in \{0,1\}^k} \in V^{\{0,1\}^k}$ such that*

$$\sum_{\omega \in F} (-1)^{|\omega|} v_{\omega} \in V_i \tag{C.4}$$

whenever $0 \leq i \leq k$ and F is a face of $\{0,1\}^k$ of dimension i .

Thus, for instance, $\mathrm{HK}^2(V)$ is the space of all quadruples $(v_{00}, v_{01}, v_{10}, v_{11})$ such that

$$\begin{aligned} v_{00}, v_{01}, v_{10}, v_{11} &\in V_0 \\ v_{00} - v_{01}, v_{00} - v_{10}, v_{01} - v_{11}, v_{10} - v_{11} &\in V_1 \\ v_{00} - v_{01} - v_{10} + v_{11} &\in V_2. \end{aligned}$$

Of course, this is equivalent to the previous description of $\mathrm{HK}^2(V)$ after a change of variables.

Proof. Let $V^{[k]}$ denote the space of all tuples $(v_\omega)_{\omega \in \{0,1\}^k} \in V^{\{0,1\}^k}$ obeying the constraints (C.4). This is clearly a subgroup of $V^{\{0,1\}^k}$. By checking the generators of $\mathrm{HK}^k(V)$, we see that $\mathrm{HK}^k(V) \subset V^{[k]}$. Now we prove the reverse inclusion $V^{[k]} \subset \mathrm{HK}^k(V)$. This claim is obvious for $k = 0$, so we may assume inductively that $k \geq 1$ and that the claim has already been proven for $k - 1$.

Let $v := (v_\omega)_{\omega \in \{0,1\}^k} \in V^{[k]}$. We split $v = v' + v''$, where

$$v' := (v_{\omega_1, \dots, \omega_{k-1}, 0})_{\omega \in \{0,1\}^k}$$

and $v'' := v - v'$. It is easy to see from (C.4), that v' and v'' lie in $V^{[k]}$. The tuple

$$(v_{\omega_1, \dots, \omega_{k-1}, 0})_{\omega \in \{0,1\}^{k-1}}$$

lies in $V^{[k-1]}$, and hence in $\mathrm{HK}^{k-1}(V)$ by induction hypothesis. Extending each generator of $\mathrm{HK}^{k-1}(V)$ to $\mathrm{HK}^k(V)$ by the homomorphism

$$(w_{\omega_1, \dots, \omega_{k-1}})_{\omega \in \{0,1\}^{k-1}} \mapsto (w_{\omega_1, \dots, \omega_{k-1}})_{\omega \in \{0,1\}^k},$$

we then see that v' lies in $\mathrm{HK}^k(V)$. In a similar spirit, we see from (C.4) that the tuple

$$(v_{\omega_1, \dots, \omega_{k-1}, 0} - v_{\omega_1, \dots, \omega_{k-1}, 0})_{\omega \in \{0,1\}^{k-1}}$$

lies in the analogue of $V^{[k-1]}$ in which the filtration $(V_n)_{n \in \mathbb{N}}$ is replaced by the shifted filtration $(V_{n+1})_{n \in \mathbb{N}}$. By induction hypothesis, this tuple thus lies in the analogue of $\mathrm{HK}^{k-1}(V)$; extending each generator of this group to $\mathrm{HK}^k(V)$ by the homomorphism

$$(w_{\omega_1, \dots, \omega_{k-1}})_{\omega \in \{0,1\}^{k-1}} \mapsto (\omega_k w_{\omega_1, \dots, \omega_{k-1}})_{\omega \in \{0,1\}^k},$$

we see that v'' lies in $\mathrm{HK}^k(V)$. Thus v lies in $\mathrm{HK}^k(V)$, and the claim follows. \square

C.17. Equidistribution. We isolate a special class of polynomial maps:

Definition C.18 (Weak equidistribution on cubes). A polynomial map $\phi : H \rightarrow G$ from one filtered group $H = (H, H_{\mathbb{N}})$ to another $G = (G, G_{\mathbb{N}})$ is said to be *weakly equidistributed on cubes* if the maps $\mathrm{HK}^k(\phi) : \mathrm{HK}^k(H) \rightarrow \mathrm{HK}^k(G)$ are surjective for every $k \in \mathbb{N}$.

Informally, a polynomial map that is weakly equidistributed on cubes can attain every possible set of values on a cube in H , subject of course to the polynomiality requirement that this set of values must form a cube in G .

The significance of weak equidistribution for us lies in the fact that they can be used to factorise polynomial maps:

Lemma C.19 (Factorisation via weak equidistribution). *Let G, H, K be filtered groups, let $\phi : H \rightarrow G$ be a polynomial map that is weakly equidistributed on cubes, and let $\psi : G \rightarrow K$ be a map. Then ψ is a polynomial map if and only if $\psi \circ \phi$ is polynomial.*

Proof. This is immediate from Theorem C.13. \square

In practice, we will derive weak equidistribution from a stronger equidistribution property, which we formulate in the language of non-standard analysis:

Definition C.20 (Equidistribution). Let A be a non-empty limit finite set, and let B be a finite set. A limit map $f : A \rightarrow B$ is said to be *equidistributed* if one has

$$\frac{1}{|A|} |\{a \in A : f(a) = b\}| = \frac{1}{|B|} + o(1)$$

for all $b \in B$.

A finite collection of functions $f_i : A \rightarrow B_i$, $i = 1, \dots, k$ into finite sets B_i is said to be *jointly equidistributed* if the combined function $(f_1, \dots, f_k) : A \rightarrow B_1 \times \dots \times B_k$ is jointly equidistributed.

Definition C.21 (Equidistribution on cubes). A polynomial map $\phi : H \rightarrow G$ from a limit-finite filtered group $H = (H, H_{\mathbb{N}})$ to a finite filtered group $G = (G, G_{\mathbb{N}})$ is said to be *(strongly) equidistributed on cubes* if the maps $\mathrm{HK}^k(\phi) : \mathrm{HK}^k(H) \rightarrow \mathrm{HK}^k(G)$ are equidistributed for every $k \in \mathbb{N}$.

Observe that if A has unbounded cardinality, then every equidistributed limit map from A to a finite set B is automatically surjective; in particular, strong equidistribution implies weak equidistribution. As a consequence of this and Lemma C.19, we obtain

Corollary C.22 (Factorisation via strong equidistribution). *Let G, H, K be filtered groups with G and G_0 limit-finite with unbounded cardinality, and H finite. Let $\phi : H \rightarrow G$ be a polynomial map that is strongly equidistributed on cubes, and let $\psi : G \rightarrow K$ be a map. Then ψ is a polynomial map if and only if $\psi \circ \phi$ is polynomial.*

For future reference, we observe a convenient criterion for equidistribution.

Lemma C.23 (Weyl equidistribution criterion). *Let A be a non-empty limit finite set, and let B be a finite abelian group. Then a limit map $f : A \rightarrow B$ is strongly equidistributed if and only if one has*

$$\mathbb{E}_{a \in A} e(\xi(f(a))) = o(1) \tag{C.5}$$

for all non-zero characters (i.e. homomorphisms) $\xi : B \rightarrow \mathbb{T}$.

Proof. By the Fourier inversion formula, the condition (C.5) is equivalent to the bound

$$\mathbb{E}_{a \in A} F(f(a)) = \mathbb{E}_{b \in B} F(b) + o(1)$$

holding for all standard functions $F : B \rightarrow \mathbb{C}$. But this is clearly equivalent in turn to the equidistribution of f . \square

APPENDIX D. PROPERTIES OF NON-CLASSICAL POLYNOMIALS

In this appendix we prove Lemma 1.7. The arguments here are analogous to those established in the context of dynamical systems in [3].

It is convenient to introduce a ring of *formal differential operators*.

Definition D.1 (Differential operators). Let V be a finite-dimensional or limit finite-dimensional vector space. A *differential operator* on V is a formal combination (using addition and multiplication) of integers and additive derivatives Δ_h (or equivalently, the shifts T_h), thus for instance $3 - 5\Delta_h + 7\Delta_h\Delta_k + \Delta_h^3$ is a differential operator. More generally, a *formal differential operator* is a finite or infinite series $\sum_i a_i \Delta_{v_{i,1}} \dots \Delta_{v_{i,d_i}}$ where $a_i \in \mathbb{Z}$, $v_{i,1}, \dots, v_{i,d_i} \in V$, and for each d there are at most finitely many i with $d_i \leq d$. Thus for instance $\sum_{i=0}^{\infty} \Delta_h^i$ is a formal differential operator. Note that both differential operators and formal differential operators act linearly (over \mathbb{Z}) on $\text{Poly}_{\leq d}(V \rightarrow G)$ for every d and G . We say that two formal differential operators on V are *equivalent* if they act the same on every space $\text{Poly}_{\leq d}(V \rightarrow G)$. We let $\text{Diff}(V)$ denote the space of formal differential operators on V modulo this equivalence relation; this is clearly a commutative ring (note that $\Delta_h\Delta_k = \Delta_k\Delta_h$ for all h, k).

Example D.2. If $h, k \in V$, then we have the *cocycle equation*

$$\Delta_{h+k} = \Delta_h + T_h\Delta_k \quad (\text{D.1})$$

in $\text{Diff}(V)$, since we have

$$\Delta_{h+k}f(x) = \Delta_h f(x) + \Delta_k f(x+h)$$

for all $f : V \rightarrow G$ and $x, h, k \in V$. This cocycle equation can also be deduced from the group law

$$T_{h+k} = T_h T_k$$

and the identity

$$\Delta_h = T_h - 1.$$

The reason for working with formal differential operators rather than genuine differential operators is that any formal differential operator of the form $1 + D$, where D consists of *higher order terms* in the sense that it contains no constant term in its expansion, is invertible in $\text{Diff}(V)$ by formal Neumann series:

$$(1 + D)^{-1} = 1 - D + D^2 - \dots$$

To illustrate this, take $h \in V$. Since $ph = 0$, we clearly have

$$T_h^p = T_{ph} = 1.$$

Expanding $T_h = 1 + \Delta_h$ and using the binomial formula, we conclude after some rearrangement that

$$\Delta_h^p = -p\Delta_h \left(1 + \frac{p-1}{2}\Delta_h + \dots + \Delta_h^{p-2}\right).$$

The expression in parentheses can be inverted by formal Neumann series. We conclude the following fundamental fact:

Lemma D.3 (Multiplication by p). *For any $h \in V$, we have $p\Delta_h = \Delta_h^p \times I_h$ for some invertible $I_h \in \text{Diff}(V)$. Furthermore, I_h is equal to -1 plus higher order terms.*

A heuristic way to interpret this lemma is that the operation of multiplication by p resembles a differential operator of order $p-1$; dually, $\frac{1}{p}$ resembles a polynomial of degree $p-1$. (This may help explain the condition $i_1 + \dots + i_n + j(p-1) \leq d$ in (1.2).)

We now begin the proof of Lemma 1.7.

The first part of claim (i) is clear by induction on d . To prove the second part, we observe from the cocycle identity (D.1) that if $\Delta_h P, \Delta_k P$ both lie in $\text{Poly}_{\leq d-1}(V \rightarrow G)$, then $\Delta_{h+k} P$ does also, and so the second part of (i) follows from the first. (One can also deduce (i) from Proposition C.10 and induction.)

Next, we establish (ii). We begin with the one-dimensional case $n = 1$. For $0 \leq d \leq p-1$, the vector space $\text{Poly}_{\leq d}(V \rightarrow \mathbb{F})$ clearly contains the vector space spanned by the monomials $1, x, \dots, x^d$, which are linearly independent as can be seen from computing a Vandermonde determinant (or using the Newton interpolation formula). On the other hand, the differential operator $\Delta_1 : \text{Poly}_{\leq d}(V \rightarrow \mathbb{F}) \rightarrow \text{Poly}_{\leq d-1}(V \rightarrow \mathbb{F})$ has kernel equal to the constant functions \mathbb{F} , and so the dimension of $\text{Poly}_{\leq d}(V \rightarrow \mathbb{F})$ can only exceed that of $\text{Poly}_{\leq d-1}$ by 1 at most. By induction we thus see that $\text{Poly}_{\leq d}(V \rightarrow \mathbb{F})$ is equal to the $d+1$ -dimensional space spanned by $1, x, \dots, x^d$ for $0 \leq d \leq p-1$. In particular, $\text{Poly}_{\leq p-1}(V \rightarrow \mathbb{F})$ must be equal to the p -dimensional space of all functions from V to \mathbb{F} , and the claim follows.

Now we assume inductively that $n > 1$ and that (ii) has already been proven for smaller dimensions. We parameterise an element $x \in \mathbb{F}^n$ as $x = (x', x_n)$ where $x' \in \mathbb{F}^{n-1}$ and $x_n \in \mathbb{F}$. If $P \in \text{Poly}_{\leq d}(V \rightarrow \mathbb{F})$, then clearly the one-dimensional maps $x_n \mapsto P(x', x_n)$ are polynomials of degree $\leq d$ for each fixed x' . Applying the one-dimensional case of (ii), we conclude that

$$P(x', x_n) = \sum_{0 \leq i_n \leq \min(p-1, d)} P_{i_n}(x') x_n^{i_n}$$

for some functions $P_{i_n} : \mathbb{F}^{n-1} \rightarrow \mathbb{F}$ that are uniquely determined by P . Differentiating this identity i_n times in the direction of the n^{th} generator e_n of \mathbb{F}^n , and $d - i_n$ times in directions in \mathbb{F}^{n-1} , we conclude that each P_{i_n} is a polynomial of degree $\leq d - i_n$. The claim then follows from the induction hypothesis.

We skip (iii) for the moment and move on to (iv). If $h \in V$ and $P, Q : V \rightarrow R$, we have

$$T_h(PQ) = (T_h P)(T_h Q);$$

expanding $T_h = 1 + \Delta_h$, we conclude the *discrete Leibniz rule*

$$\Delta_h(PQ) = (\Delta_h P)Q + P(\Delta_h Q) + (\Delta_h P)(\Delta_h Q). \quad (\text{D.2})$$

The claim (iv) can now be easily established by an induction on $d + d'$ (noting that the claim is trivial if d or d' is negative). (Alternatively, one can deduce (iv) from [24, Example B.18].)

We remark that one should view the final term $(\Delta_h P)(\Delta_h Q)$ in (D.2) to be a lower order error term, so that (D.2) becomes a perturbation of the classical Leibniz rule $D(PQ) = (DP)Q + P(DQ)$ for derivations D .

Now we prove part of (v). If $P \in \text{Poly}_{\leq d}(V \rightarrow G)$ for some $d \geq p-1$, then

$$\Delta_{h_1}^p \Delta_{h_2} \dots \Delta_{h_{d-p+2}} P = 0$$

for any $h_1, \dots, h_{d-p+2} \in V$, and thus by Lemma D.3

$$\Delta_{h_1} \Delta_{h_2} \dots \Delta_{h_{d-p+2}} pP = 0.$$

We conclude that $pP \in \text{Poly}_{\leq d-p+1}(V \rightarrow G)$. We conclude that for any integer d , the map $p : P \mapsto pP$ maps $\text{Poly}_{\leq d}(V \rightarrow G)$ to $\text{Poly}_{\leq \max(d-p+1, 0)}(V \rightarrow G)$. This proves everything in (v) except for the assertion that this map $P \mapsto pP$ is surjective.

Now we return to (iii). The fact that every expression P of the form (1.2) is a polynomial of degree $\leq d$, and vice versa follows from the special case of Proposition 10.4 when all the initial degrees D_i are equal to 1. (This argument is non-circular, because Lemma 1.7(iii) is not used in the proof of Proposition 10.4. Another proof of this part of Lemma 1.7(iii) can be found in [41, §1.12].

Now we establish the uniqueness claim in (iii). The claim is trivial for $d = 0$, so suppose inductively that $d \geq 1$ and that uniqueness has already been established for smaller values of d . Since $\alpha = P(0)$ from (1.2) we see that the α are unique and can thus be subtracted away. Applying the uniqueness claim to the lower-degree polynomial pP , we see that all the coefficients in (1.2) with $j \geq 1$ are unique. Subtracting off these terms also, we are left with a classical polynomial expansion (1.1), and the claim follows from (ii). This concludes the proof of (iii).

Now that we have (iii), the surjectivity claim of (v) is immediate, since one simply replaces all the p^{j+1} denominators in (1.2) by p^{j+2} , and replaces α with a p^{th} root α' as in the proof of (iii). This completes the proof of (v).

The claim (vi) follows immediately from (v) and an induction on d .

APPENDIX E. ON EXACT ROOTS IN DYNAMICAL SYSTEMS

In [3], [43], the inverse conjecture for the Gowers norms were attacked via an ergodic theory approach, based on a structural analysis of \mathbb{F}^ω -systems. These systems consisted of a probability space (X, \mathcal{B}, μ) , together with a measure-preserving action $(T_g)_{g \in \mathbb{F}^\omega}$ of the infinite vector space $\mathbb{F}^\omega := \bigcup_{n=1}^\infty \mathbb{F}^n$ (where we nest \mathbb{F}^n inside \mathbb{F}^{n+1} in the obvious manner).

A polynomial of degree $\leq d$ on such a system is a measurable function $P : X \rightarrow \mathbb{T}$ such that $\Delta_{h_1} \dots \Delta_{h_{d+1}} P = 0$ a.e. for all $h_1, \dots, h_{d+1} \in \mathbb{F}^\omega$, where $\Delta_h f := f \circ T_h - 1$. The analogue of the exact roots property from Remark 1.9 is then

Claim E.1. *Let $X = (X, \mathcal{B}, \mu, (T_g)_{g \in \mathbb{F}^\omega})$ be a \mathbb{F}^ω -system, let $d \geq 0$ be an integer, and let $P : X \rightarrow \mathbb{T}$ be a polynomial of degree $\leq d$. Then there exists a polynomial $Q : X \rightarrow \mathbb{T}$ of degree $\leq d + p - 1$ with $pQ = P$.*

Using Lemma 1.7(v), it is not difficult to verify this claim when X is finite; the claim is also easy when $d = 1$, as one can then (up to a constant) express $P = \iota(\tilde{P})$ for some linear polynomial $\tilde{P} : X \rightarrow \mathbb{F}$, and one can verify that the polynomial $Q := e(|\tilde{P}|/p^2)$ will

have the desired properties (note from Lemma 1.7(iii) that the map $x \mapsto |x|/p^2 \bmod 1$ has degree $\leq p$ on \mathbb{F}). Unfortunately, the claim fails in general. For instance, we have

Proposition E.2. *Claim E.1 is false when $p = 2$ and $d = 2$.*

The purpose of this appendix is to prove this proposition, which explains why we were unable to use the ergodic theory method from [43], [3] to establish the main results in this paper.

The reason that the exact roots property holds in the finitary setting but not in the ergodic setting can be explained as follows. In the finite setting \mathbb{F}_2^n , all functions are considered to be measurable; but in the ergodic setting, only a limited number of functions are measurable. For instance, one may be working in a system generated by a single function f and its shifts $T_h f$, so that every measurable function in the system can be approximated to arbitrary accuracy by a finite combination of shifts $T_h f$ of these functions.

Now consider the finitary quadratic function $P := \frac{L}{4} \bmod 1$ on \mathbb{F}_2^n , which is the analogue of the function $\iota_4(t)$ considered above. This function has a degree ≤ 3 root, namely the function $Q := \frac{L}{8} \bmod 1$. However, this polynomial Q is not “measurable” in the system generated by P , in the sense that one cannot express (even approximately) Q as a function of a bounded number of shifts $\Delta_h P$ of P ; indeed, one can formulate this precisely and then deduce this from the arguments in the proof of Proposition E.2 given below, combined with the Furstenberg correspondence principle, but we will not do this here.

On the other hand, if we allow ourselves the freedom to *extend* the system X to a larger one, then it appears that one can recover the exact roots property. For instance, in the system $X = \mathbb{F}^{\mathbb{N}} \times_{\rho} \mathbb{Z}/4\mathbb{Z}$ defined below, we may extend this system to the system $Y := \mathbb{F}^{\mathbb{N}} \times_{\rho_8} \mathbb{Z}/8\mathbb{Z}$, where the cocycle ρ_8 is defined exactly as with $\rho = \rho_4$ but using the modulus 8 rather than 4. Letting t_8 be the vertical coordinate function of Y , the degree ≤ 2 polynomial $\iota_4(t)$ in X then lifts to $2\iota_8(t_8)$ in Y , where $\iota_8 : \mathbb{Z}/8\mathbb{Z} \rightarrow \{0, 1, \dots, 7\}$ is the obvious map. This has an obvious root that is of degree ≤ 3 , namely $\iota_8(t_8)$; this function is the ergodic analogue of the finitary function $Q = \frac{L}{8} \bmod 1$ considered earlier. It is likely that this phenomenon generalises, in that Claim E.1 becomes true again if we allow Q to take values in an extension of P , but we will not pursue this statement here. (Note though that one can use extensions to simplify the proofs of various multiple recurrence and convergence results in ergodic theory; see for instance [2].)

In principle, this weakened form of Claim E.1, in which the root takes values in an extension, may possibly be used to extend the ergodic theory arguments in [3] to the low characteristic setting, and in particular to recover the ergodic version of the inverse conjecture for the Gowers norms for \mathbb{F}^{ω} (see [3], [43]) in that setting. However, we were not able to achieve this, as the inductive arguments in [3] rely on reducing the system and are thus not compatible with taking extensions.

We now begin the formal proof of Proposition E.2. Let $\mathbb{F} := \mathbb{F}_2$. Consider the Cantor space $\mathbb{F}^{\mathbb{N}} = \prod_{n \in \mathbb{N}} \mathbb{F}^{\mathbb{N}}$ (with the product σ -algebra and the uniform (Bernoulli) probability measure), which has an obvious action of \mathbb{F}^{ω} . If we let $x_i : \mathbb{F}^{\mathbb{N}} \rightarrow \mathbb{F}$, $i \in \mathbb{N}$ be the coordinate functions, we thus have

$$\Delta_{e_i} x_j = \delta_{ij}$$

for $i, j \in \mathbb{N}$, where e_1, e_2, \dots are the generators of \mathbb{F}^{ω} and δ_{ij} is the Kronecker delta. In particular, the functions $\iota(x_j) : X \rightarrow \mathbb{T}$ are polynomials of degree ≤ 1 .

We define a cocycle $\rho = \rho_4 : \mathbb{F}^{\omega} \times \mathbb{F}^{\mathbb{N}} \rightarrow \mathbb{Z}/4\mathbb{Z}$ taking values in the cyclic group $\mathbb{Z}/4\mathbb{Z}$ by the formula

$$\rho\left(\sum_{i \in A} e_i, x\right) := \sum_{i \in A} (1 - 2|x_i|) \bmod 4 \quad (\text{E.1})$$

for any finite set A . One easily verifies the *cocycle equation*

$$\rho(h + k, x) = \rho(h, x) + \rho(k, T_h x)$$

for any $h, k \in \mathbb{F}^{\omega}$. We can therefore build the cocycle extension $X := \mathbb{F}^{\mathbb{N}} \times_{\rho} \mathbb{Z}/4\mathbb{Z}$ of $\mathbb{F}^{\mathbb{N}}$, defined as the space of pairs (x, t) with $x \in \mathbb{F}^{\mathbb{N}}$ and $t \in \mathbb{Z}/4\mathbb{Z}$ (with the product probability measure) endowed with the shift

$$T_h(x, t) := (T_h x, t + \rho(h, x)).$$

This can be easily verified to be a \mathbb{F}^{ω} -system. If we let $\iota_4 : \mathbb{Z}/4\mathbb{Z} \rightarrow \mathbb{T}$ be the map $\iota_4(i) := i/4$, and $t : (x, t) \mapsto t$ be the coordinate function, then the function $\iota_4(t)$ has derivatives

$$\Delta_h \iota_4(t) = \iota_4(\rho(h, \cdot)),$$

and thus by (E.1), $\iota_4(t)$ is a polynomial of degree ≤ 2 . Similarly, $2\iota_4(t)$ is a polynomial of degree ≤ 1 .

If Claim E.1 was true, then there would exist a polynomial $Q : X \rightarrow \mathbb{T}$ of degree ≤ 3 such that $2Q = \iota_4(t)$. To show that this is not possible, we need to classify all the polynomials of degree ≤ 3 .

We begin with the polynomials of degree ≤ 0 , which (up to measure zero errors) are simply the \mathbb{F}^{ω} -invariant functions $P : X \rightarrow \mathbb{T}$. We claim that X is ergodic, so that the only invariant functions are the constants (up to measure zero errors). It suffices to show that every invariant set E in X has zero measure or full measure. Given any ε , we can approximate E to an error of measure ε by a set E_{ε} which depends on only a finite number x_1, \dots, x_n of the base coordinate functions, together with the vertical coordinate t . The set E_{ε} is then invariant up to errors of measure $O(\varepsilon)$. Inspecting the action of the shift $T_{e_{n+1}}$ on E_{ε} , one then easily concludes that E_{ε} must differ by an error of $O(\varepsilon)$ from a set which does not depend on the vertical coordinate. Taking $\varepsilon \rightarrow 0$, we conclude that E is (up to measure zero errors) independent of the vertical coordinate, and thus descends to an invariant subset of $\mathbb{F}^{\mathbb{N}}$. But it is standard from the theory of Bernoulli systems that such sets have either zero measure or full measure.

Now we classify the polynomials of degree ≤ 1 . If P has degree ≤ 1 , then for each e_i , $\Delta_{e_i} P$ is degree ≤ 0 and hence constant. Since $\Delta_{2e_i} P$ is necessarily trivial, we conclude

from the cocycle equation $\Delta_{h+k} = \Delta_h + T_h \Delta_k$ with $h = k = e_i$ that $2\Delta_{e_i}P = 0$. Thus we can find coefficients $c_i \in \mathbb{F}$ such that $\Delta_{e_i}P = c_i$ for all i . Also, since P is measurable, it differs by an error of ε (in measure) from a function which depends on only finitely many of the x_1, \dots, x_n and t . Note that if $n_1 > n_2 > n$ and (x, t) are such that $x_{n_1} \neq x_{n_2}$, then the shift $T_{e_{n_1} - e_{n_2}}$ does not affect the x_1, \dots, x_n, t coefficients. On the other hand, $T_{e_{n_1} - e_{n_2}}P = P + c_{n_1} - c_{n_2}$. and thus (if ε is small enough) we have $c_{n_1} = c$ independent of n_1 for $n_1 > n$. We now see that P differs from the degree ≤ 1 polynomial

$$\sum_{i=1}^n (c_i - c)\iota(x_i) + 2\iota_4(t)$$

by a degree ≤ 0 polynomial, which is thus constant. Thus all degree ≤ 1 polynomials take the form

$$P = \sum_i c_i \iota(x_i) + 2d\iota_4(t) + \alpha$$

where $\alpha \in \mathbb{T}$, $d \in \mathbb{Z}$, and at most finitely many of the $c_i \in \mathbb{Z}$ are non-zero.

To classify polynomials of higher degree we employ the method of *vertical differentiation*, which is used extensively in the ergodic theory literature (see for instance [26], [46], [3]); for this simple example we use a very concrete instance of this method here. We define the vertical derivatives $\Delta_s P$ of a function $P : X \rightarrow \mathbb{T}$ for any $s \in \mathbb{Z}/4\mathbb{Z}$ by the formula

$$\Delta_s P(x, t) := P(x, t + s) - P(x, t).$$

Observe that these operators commute with themselves and with the \mathbb{F}^ω action, and in particular commute with the ordinary derivatives Δ_h .

The key observation is that Δ_2 behaves like a differential operator of order two:

Lemma E.3. *Let $s = 1, 2$. If $P : X \rightarrow \mathbb{T}$ has degree $\leq d$ for some integer d , then $\Delta_s P$ has degree $\leq d - s$.*

Proof. By repeated differentiation it suffices to verify this when $d = s - 1$. But this follows from the classification of polynomials of degree ≤ 0 and degree ≤ 1 that has already been established. \square

We can now classify polynomials of degree ≤ 2 (cf. Lemma 1.7(iii)):

Lemma E.4 (Classification of quadratics). *Let $P : X \rightarrow \mathbb{T}$ be of degree ≤ 2 . Then we can write*

$$P = \sum_{i < j} c_{ij} \frac{|x_i||x_j|}{2} + \sum_i c_i \frac{|x_i|}{4} + \sum_i d_i \frac{|x_i||S_1|}{2} + d \frac{|S_1|}{4} + e \frac{|t|}{4} + \alpha \pmod{1} \quad (\text{E.2})$$

where c_{ij}, c_i, d_i, d, e are integers, of which only finitely many are non-zero, $\alpha \in \mathbb{T}$, and $S_1 : X \rightarrow \mathbb{F}$ is the function $S_1 := t \pmod{2}$, and $x \mapsto |x|$ for $x \in \mathbb{Z}/4\mathbb{Z}$ is the obvious map to the fundamental domain $\{0, 1, 2, 3\}$.

Proof. By a computation, one verifies that all the expressions on the right-hand side of (E.2) are indeed polynomials of degree ≤ 2 .

From Lemma E.3, $\Delta_2 P$ is constant. Using the cocycle identity

$$\Delta_2 P(x, t) + \Delta_2 P(x, t + 2) = 0 \quad (\text{E.3})$$

we see that this constant is either 0 or $1/2$. In the latter case, we can subtract off $|t|/4$ to reduce to the former case (noting that $\Delta_2 |t|/4 = 1/2$); so we may assume without loss of generality that $\Delta_2 P = 0$, thus P descends to the reduced system $\mathbb{F}^\omega \times_{\rho \bmod 2} \mathbb{Z}/2\mathbb{Z}$.

From Lemma E.3 again, $\Delta_1 P$ is linear, and thus takes the form

$$\Delta_1 P(x, t) = \sum_i d_i \frac{|x_i|}{2} + \beta$$

for some integers d_i (of which only finitely many are non-zero) and $\beta \in \mathbb{T}$. With the cocycle identity

$$\Delta_1 P(x, t) + \Delta_1 P(x, t + 1) = \Delta_2 P(x, t) = 0 \quad (\text{E.4})$$

we see that β is either 0 or $1/2$, thus $\beta = d/2$ for some integer d . If we then subtract off $\sum_i d_i \frac{|x_i||S_1|}{2} + d \frac{|S_1|}{4}$ from P (noting that $\Delta_1 S_1 = 1$) we can reduce to the case $\Delta_1 P = 0$, thus P now descends to a function of $\mathbb{F}^\mathbb{N}$.

For any ε , we may approximate P in measure to error ε by a function depending only on finitely many x_1, \dots, x_n of the coefficients. For any $n' > n$, the linear polynomial $\Delta_{e_{n'}} P$ is then within $O(\varepsilon)$ of zero in measure, and is thus constant (by the classification of linear polynomials); using the cocycle identity $0 = \Delta_{e_{n'}} P + T_{e_{n'}} \Delta_{e_{n'}} P$ we see that $\Delta_{e_{n'}} P$ is in fact identically zero, thus P in fact descends to a function of just a finite number of coordinates x_1, \dots, x_n . The claim now follows from Lemma 1.7(ii). \square

In a similar vein, we can classify cubics:

Lemma E.5 (Classification of cubics). *Let $P : X \rightarrow \mathbb{T}$ be of degree ≤ 3 . Then P is an integer linear combination of a finite number of the following functions:*

- (i) Constants $\alpha \in \mathbb{T}$;
- (ii) $|x_i||x_j||x_k|/2$ for natural numbers $i < j < k$;
- (iii) $|x_i||x_j|/4$ for natural numbers $i < j$;
- (iv) $|x_i|/8$ for a natural number i ;
- (v) $|x_i||x_j||S_1|/2$ for natural numbers $i < j$;
- (vi) $|x_i||S_1|/4$ for a natural number i ;
- (vii) $|S_1|/8$ for a natural number i ;
- (viii) $|x_i||S_2|/2$ for a natural number i ;
- (ix) $|S_1||S_2|/2$;
- (x) $|S_2|/2$.

Here $S_2 : X \rightarrow \mathbb{F}$ is the function such that $S_2(x, t) = 1$ when $t = 2, 3 \bmod 4$ and $S_2(x, t) = 0$ otherwise.

Remark E.6. The polynomials S_1, S_2 can be viewed as the ergodic limit as $n \rightarrow \infty$ (using the Furstenberg correspondence principle) of the symmetric polynomials S_1, S_2 considered in Example 1.16, where the coordinate functions x_i correspond to the usual

coordinate functions on \mathbb{F}_2^n , and t corresponds to $L \bmod 4$. The cubic polynomial $\iota_4(t) = |t|/4$ can be expressed as $|t|/4 = |S_2|/2 + |S_1|/4$; cf. (1.5).

Proof. By a (somewhat tedious) computation we see that all the above functions are polynomials of degree ≤ 3 (and that S_2 is a polynomial of degree ≤ 2).

From Lemma E.3, $\Delta_2 P$ is of degree ≤ 1 , so we can express

$$\Delta_2 P = \sum_i c_i |x_i|/2 + c |S_1|/2 + \alpha$$

for some integers c_i, c (only finitely many of which are non-zero) and $\alpha \in \mathbb{T}$. Using (E.3) we can write $\alpha = d/2$ for some integer d . By subtracting $\sum_i c_i |x_i| |S_2|/2 + c |S_1| |S_2|/2 + d |S_2|/2$ from P we may thus assume that $\Delta_2 P = 0$.

Next, $\Delta_1 P$ is of degree ≤ 2 , and is annihilated by Δ_2 , so by Lemma E.4, we have

$$\Delta_1 P = \sum_{i < j} c_{ij} \frac{|x_i| |x_j|}{2} + \sum_i c_i \frac{|x_i|}{4} + \sum_i d_i \frac{|x_i| |S_1|}{2} + d \frac{|S_1|}{4} + \alpha \bmod 1$$

for some integers c_{ij}, c_i, d_i, d (only finitely many of which are non-zero) and $\alpha \in \mathbb{T}$. Using (E.4), we conclude that the $\frac{c_i + d_i}{2}$ must vanish, and that $\frac{d}{4} + 2\alpha = 0$; thus we can simplify the above expression to

$$\Delta_1 P = \sum_{i < j} c_{ij} \frac{|x_i| |x_j|}{2} + \sum_i d_i \left(\frac{|x_i| |S_1|}{2} - \frac{|x_i|}{4} \right) + d \left(\frac{|S_1|}{4} - \frac{1}{8} \right) + \frac{e}{2} \bmod 1$$

for some integer e . If we then subtract off the cubic polynomial

$$\sum_{i < j} c_{ij} \frac{|x_i| |x_j| |S_1|}{2} + \sum_i d_i \frac{|x_i| |S_1|}{4} + d \frac{|S_1|}{8} + \frac{e |S_1|}{2}$$

from P , we can reduce to the case $\Delta_1 P = 0$, thus P descends to $\mathbb{F}^{\mathbb{N}}$. Arguing as in Lemma E.4, we conclude that P is a function of finitely many coordinates x_1, \dots, x_n , and the claim follows from Lemma 1.7(iii). \square

From the above lemma, we see that if P is a polynomial of degree ≤ 3 , then $2P$ is an integer linear combination of the following types of functions:

- (i) Constants $\alpha \in \mathbb{T}$;
- (ii) $|x_i| |x_j|/2$ for natural numbers $i < j$;
- (iii) $|x_i|/4$ for a natural number i ;
- (iv) $|x_i| |S_1|/2$ for a natural number i ;
- (v) $|S_1|/4$ for a natural number i .

In particular, $\Delta_2(2P)$ must vanish. On the other hand, $\Delta_2 \iota_4(t) = \frac{1}{2} \neq 0$. Thus $\iota_4(t)$ is not of the form $2P$ for a polynomial of degree ≤ 3 , thus establishing Proposition E.2.

Remark E.7. One can avoid the full classification of polynomials of degree ≤ 3 in proving the above proposition. Indeed, if P is cubic, then from Lemma E.3 $\Delta_2 \Delta_2 P = 0$, and then by (E.3) $\Delta_2(2P) = 0$, and we can conclude as above.

REFERENCES

- [1] N. Alon, T. Kaufman, M. Krivelevich, S. Litsyn and D. Ron, *Testing flow-degree polynomials over $GF(2)$* , RANDOM-APPROX 2003, 188–199. Also: Testing Reed-Muller codes, IEEE Transactions on Information Theory 51 (2005), 4032–4039.
- [2] T. Austin, *On the norm convergence of nonconventional ergodic averages*, preprint.
- [3] V. Bergelson, T. Tao and T. Ziegler, *An inverse theorem for the uniformity seminorms associated with the action of \mathbb{F}_p^∞* , preprint.
- [4] V. Bergelson, A. Leibman, R. McCutcheon, *Polynomial Szemerédi theorems for countable modules over integral domains and finite fields*, J. Anal. Math. **95** (2005), 243–296.
- [5] A. Bogdanov, E. Viola, *Pseudorandom bits for polynomials*, Proc. of FOCS 2007, 41–51.
- [6] O. Camarena, B. Szegedy, *Nilspaces, nilmanifolds and their morphisms*, preprint.
- [7] N. Frantzikinakis, B. Host, B. Kra, *Multiple recurrence and convergence for sequences related to the prime numbers*, preprint.
- [8] H. Furstenberg, *Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions*, J. Analyse Math. **31** (1977), 204–256.
- [9] H. Furstenberg, Y. Katznelson, *A density version of the Hales-Jewett theorem*, J. Anal. Math. **57** (1991), 64–119.
- [10] K. Gödel, *Consistency of the axiom of choice and of the generalized continuum-hypothesis with the axioms of set theory*, Proc. Nat. Acad. Sci. **24** (1938), 556–557.
- [11] T. Gowers, *A new proof of Szemerédi’s theorem for arithmetic progressions of length four*, Geom. Func. Anal. **8** (1998), 529–551.
- [12] T. Gowers, *A new proof of Szemerédi’s theorem*, Geom. Func. Anal., **11** (2001), 465–588.
- [13] T. Gowers, *Decompositions, approximate structure, transference, and the Hahn-Banach theorem*, preprint.
- [14] T. Gowers, J. Wolf, *The true complexity of a system of linear equations*, preprint.
- [15] T. Gowers, J. Wolf, *Linear forms and quadratic uniformity for functions on \mathbb{F}_p^n* , preprint.
- [16] T. Gowers, J. Wolf, *Linear forms and higher-degree uniformity for functions on \mathbb{F}_p^n* , preprint.
- [17] B. Green, T. Tao, *The primes contain arbitrarily long arithmetic progressions*, Annals of Math., Annals of Math. **167** (2008), 481–547.
- [18] B. Green, T. Tao, *An inverse theorem for the Gowers $U^3(G)$ norm*, Proc. Edin. Math. Soc. **51** (2008), 73–153.
- [19] B. Green, T. Tao, *Linear equations in primes*, Annals of Math., to appear.
- [20] B. Green, T. Tao, *New bounds for Szemerédi’s Theorem, I: Progressions of length 4 in finite field geometries*, Proc. Lond. Math. Soc. **98** (2009), 365–392.
- [21] B. Green, T. Tao, *The quantitative behaviour of polynomial orbits on nilmanifolds*, Annals of Math., to appear.
- [22] B. Green, T. Tao, *The distribution of polynomials over finite fields, with applications to the Gowers norms*, preprint.
- [23] B. Green, T. Tao, *An arithmetic regularity lemma, an associated counting lemma, and applications*, preprint.
- [24] B. Green, T. Tao, T. Ziegler, *An inverse theorem for the Gowers $U^{s+1}[N]$ norm*, preprint.
- [25] P. Hall, *A contribution to the theory of groups of prime-power order*, Proc. London Math. Soc. (2), **36** (1933), 29–95.
- [26] B. Host, B. Kra, *Nonconventional ergodic averages and nilmanifolds*, Ann. of Math. (2) **161** (2005), no. 1, 397–488.
- [27] B. Host, B. Kra, *Parallelepipedes, Nilpotent Groups, and Gowers Norms*, preprint.
- [28] T. Kaufman, S. Lovett, *Worst Case to Average Case Reductions for Polynomials*, FOCS 2008: 166–175.

- [29] M. Lazard, *Sur certaines suites d'éléments dans les groupes libres et leurs extensions*, C. R. Acad. Sci. Paris **236**, (1953), 36-38.
- [30] A. Leibman, *Polynomial sequences in groups*, Journal of Algebra **201** (1998), 189–206.
- [31] A. Leibman, *Polynomial mappings of groups*, Israel J. Math. **129** (2002), 29–60.
- [32] P. A. Loeb, *Conversion from nonstandard to standard measure spaces and applications in probability theory*, Trans. Amer. Math. Soc. **211** (1975), pp. 113–122.
- [33] S. Lovett, R. Meshulam, A. Samorodnitsky, *Inverse conjecture for the Gowers norm is false*, STOC 2008.
- [34] J. Petresco, *Sur les commutateurs*, Math. Z. **61** (1954), 348-356.
- [35] A. Samorodnitsky, *Low-degree tests at large distances*, STOC 2007.
- [36] M. Sudan, L. Trevisan, S. Vadhan, *Pseudorandom generators without the XOR lemma*, Special issue on the Fourteenth Annual IEEE Conference on Computational Complexity (Atlanta, GA, 1999).
- [37] B. Szegedy, *Gowers norms, regularization and limits of functions on abelian groups*, preprint.
- [38] B. Szegedy, *Structure of finite nilspaces and inverse theorems for the Gowers norms in bounded exponent groups*, preprint.
- [39] T. Tao, *A quantitative ergodic theory proof of Szemerédi's theorem*, Electron. J. Combin. **13** (2006) 1 No. 99, 1–49.
- [40] T. Tao, *Structure and randomness in combinatorics*, Proceedings of the 48th annual symposium on Foundations of Computer Science (FOCS) 2007, 3–18.
- [41] T. Tao, *Poincaré's legacies*, Vol I., American Mathematical Society, 2009.
- [42] T. Tao, V. Vu, *Additive Combinatorics*, Cambridge Univ. Press, 2006.
- [43] T. Tao, T. Ziegler, *The inverse conjecture for the Gowers norm over finite fields via the correspondence principle*, preprint.
- [44] H. Towsner, *A Correspondence Principle for the Gowers Norms*, preprint.
- [45] P. Varnavides, *On certain sets of positive density*, J. London Math. Soc. **34** (1959) 358–360.
- [46] T. Ziegler, *Universal characteristic factors and Furstenberg averages*, J. Amer. Math. Soc. **20** (2007), 53–97.

UCLA DEPARTMENT OF MATHEMATICS, LOS ANGELES, CA 90095-1596.

E-mail address: tao@math.ucla.edu

DEPARTMENT OF MATHEMATICS, TECHNION, HAIFA, ISRAEL 32000.

E-mail address: tamarzr@tx.technion.ac.il